IDA

INSTITUTE FOR DEFENSE ANALYSES

Indicators of Training Readiness

Jesse Orlansky
Colin P. Hammon
Stanley A. Horowitz

March 1997

IDA Paper P-3283

Log: H 96-004085

This work was conducted under contract DASW01 94 C 0054, Task T-L2-1371, for the Office of the Under Secretary of Defense (Personnel and Readiness). The publication of this IDA document does not indicate endorsement by the Department of Defense, nor should the contents be construed as reflecting the official position of that Agency.

INSTITUTE FOR DEFENSE ANALYSES

IDA Paper P-3283

Indicators of Training Readiness

Jesse Orlansky
Colin P. Hammon
Stanley A. Horowitz

March 1997

PREFACE

This study was performed for the Under Secretary of Defense (Personnel and Readiness) on a task entitled, "Indicators of Joint Readiness." Technical cognizance for this task was assigned to John Walsh and Daniel Gardner.

ACKNOWLEDGMENTS

We are grateful to Carol A. Johnson of the Defense Manpower Data Center, and Dexter Fletcher, who reviewed a draft of this paper and provided many helpful suggestions. Lee Ann Gehman guided the preparation of the manuscript with wisdom and professional competence.

CONTENTS

TABLES

FIGURES

SUMMARY

A. BACKGROUND

The purpose of this paper is to review training readiness indicators used now by the Services, Joint Staff, and the Office of the Secretary of Defense and to propose indicators that could improve current methods of reporting training readiness. The current method of reporting, called the Status of Resources and Training System, or SORTS, has been criticized because it includes various subjective, rather than objective assessments of training readiness and because it focuses on current readiness and does not provide estimates of training readiness in the near future.

Readiness is, as the Deputy Under Secretary of Defense for Readiness describes it:

> the general ability of forces to arrive where they are needed, on time and prepared to effectively carry out assigned mission objectives for which they were designed. The ability of units to be ready on time to carry out their missions, in turn, is a function of having the equipment, supplies, logistics and experienced people with the skills to accomplish assigned tasks. (Finch 1996)

The Services, the Secretary of Defense, the Chairman of the Joint Chiefs of Staff, and the regional Commanders-in-Chief need and rely upon regular reports of the training readiness of the forces. These reports are used to assign or shift forces on the basis of their current readiness to various contingencies as well as to require additional training where needed. The Services regularly provide many reports on their readiness with respect to training, equipment, logistics, and the like. Here, we are concerned with improving the quality of reports dealing primarily with training readiness, without detracting from the importance of reports on other factors, such as equipment and logistics, needed for success in combat.

B. APPROACH

We based our research on the hypothesis that there are three types of training readiness indicators, each with particular utility for some part of the training community. These follow a typical production function with phases of Input-Process-Output as shown below. The terms "Programmed/Actual" and "Current/Future" in each column are intended

to note that, for analytical purposes, it is important to distinguish between what is planned and what actually happened as well as between current readiness and future readiness.

Input	Process	Output
Resources Costs	Training Courses Events Exercises OPTEMPO Accomplishments	Performance Measures of effectiveness Exercise outcomes Performance to standards Training readiness
Programmed/Actual Current and Future	Programmed/Actual Current and Future	Programmed/Actual Current and Future

Using this framework, we conducted an extensive literature review and a series of interviews at the Major Command, wing, battalion, and squadron levels of the military Services.

C. FINDINGS

1. Literature Review

The literature provides information on alternative indicators of readiness that have been proposed and/or analyzed in the past. We were looking particularly for evidence that performance measures (output) could be related to measures of training accomplishment (process). We were also looking for indicators of joint readiness. Several initiatives at the Joint level, and at least one at the service level, began with the intent of improving our current knowledge of the actual readiness of forces at the Joint level, but these are all in the beginning stages. It is also clear that budgets for the cost of collective training and of exercises are still unknown. This derives partly from the DoD accounting system, but perhaps more from the practice of diverting money budgeted for training to other purposes, such as contingencies, drug interdiction, and unfunded base support.

a. Proposed Training Readiness Indicators

The Defense Science Board (DSB, 1994), the General Accounting Office (GAO, March 1995) and the Center for Naval Analyses (Robinson, et al., 1996) have proposed or cited many Service indicators that are considered essential to monitoring readiness, or are considered critical to preventing a hollow force. Most of these reports are based on expert judgment of panels of retired and active senior officers. These indicators include:

Personnel (quality, strength, turbulence, morale, retention)

Training (weapons system proficiency, required and specialty)

Equipment (amount available, condition, maintenance)

Supply

Operating Tempo

Commitments and deployments

Funding

Accident rates

Although these indicators appear likely, at least intuitively, to influence readiness, no analysis was provided to show that variations in any of them are consistently related to variations in readiness. Such work needs to be done before one should conclude that adding any of these indicators would improve our ability to evaluate current or predict future readiness.

b. Training Resource Models

The literature relating training accomplishment to performance goes back several decades and contains valuable information. The studies are of two general types: those which relate resources to training accomplishment in SORTS or its predecessor, UNITREP, and those which relate training accomplishment (flying hours, vehicle miles driven, ship days underway, exercises completed, percent of air crews combat ready, etc.) to exercise or combat performance (bombing, missile or gunnery accuracy, operational inspections, combat deaths, etc.). A perhaps surprising finding was that, in general, SORTS has been a fairly reliable indicator of readiness. Although limited in number and not necessarily an endorsement of SORTS, studies which relate performance indicators to C-1 ratings in SORTS clearly show that certain SORTS indicators have demonstrable validity.

One must concur with suggestions to improve SORTS that call for

- More objective performance measures
- Use of independent performance evaluators (i.e., from outside the unit being evaluated)
- Surprise evaluations
- Combat mission-oriented performance measures.

An important limitation seems to be that SORTS does not take into account specific mission tasks or operational environments related to missions short of full-scale wartime commitment.

Among studies which relate training accomplishment to performance we found 29 analyses that show that certain indicators of training are valid (i.e., statistically significant) predictors of combat capabilities based on data collected in field exercises or in actual combat. These are summarized in the following list:

Flying hours	Navy, Marine Corps, Air Force
Vehicle miles	Army
Ship days underway	Navy
Personnel turnover	Navy
Length of time of officers in command in combat	Army

2. Service Interviews

We found that the Services now collect certain information that is objective in nature and that is a potential indicator of training readiness. These indicators are generally available at major command levels and their use at Service Headquarters or in the Department of Defense and the Joint Staff would not impose new data collection efforts. Service databases were evaluated according to the following criteria:

- They should be easy to understand, particularly at upper levels of the Services, OSD, and the Joint Staff.
- They should not add to the reporting burden of individual units.
- There is no need to identify individual units in reports to OSD and the Joint Staff.
- The data should reflect training accomplishment and performance in separate categories.
- Training readiness indicators should be objectively determined. They should be measured against a standard rather than reflect the relative standing of like units. In addition to objective data, human judgment is useful in the determination of whether training accomplishment or performance is up to standard (as long as well-defined and well-known standards and criteria exist for comparison with the standards).
- Indicators should be discriminating and consistent. To be useful as indicators of training readiness, databases should discriminate among various levels of readiness and do so consistently.

- Training readiness indicators should provide broad coverage across services and kinds of units. They should be updated often enough to reflect current conditions.

Archiving such data provides the basis for observing trends, establishing limits beyond which various changes can be noted as significant, and providing the information needed to relate changes in resources to changes in readiness.

Recommended indicators are summarized below:

Type of indicator	Indicator	Service
Demonstrated training performance	percent of crews or platoons qualified	Army Marine Corps
	percent of submarine Training Readiness Examination above/below average	Navy subsurface
	percent of Operational Readiness Inspection excellent or outstanding	Air Force
	percent of tasks trained to standard	Army Marine Corps
Training accomplishment	percent of mission essential tasks trained	Army Marine Corps
	percent training accomplished by primary mission areas	Navy
	percent of training accomplished (percent crews combat ready)	Navy aviation USMC aviation
	percent Graduated Combat Capability level B or A	Air Force
	percent participation in Combat Training Centers/Combat Arms Exercises	Army Marine Corps

Of the systems used by the Services to compile and report training readiness information, the Navy's Type Commander Readiness Management System (TRMS) has the most comprehensive database, and its software is best suited for examining the capability of training readiness indicators. It appears to have the potential to be used as a system that could track the training readiness of Joint forces. It contains modules that provide data on equipment casualty status, training readiness, personnel, inspections, combat systems, ship readiness, and an executive summary. Because it covers individual missions, it provides an excellent basis for developing training readiness indicators at the Joint level.

D. RECOMMENDATIONS

Information on training readiness should be organized on the basis of a production model, with phases of input, process, and output, i.e., a resources-to-readiness paradigm. These data are already available in the Status of Resources and Training System (SORTS), the Navy Training Readiness Management System (TRMS), and closely related personnel and equipment status databases. The following steps should be taken:

1. Analyze data to identify short term and long term trends, including noise, i.e., short term, non-significant variations.
2. Where trends are observed, identify the time delays between inputs and outputs, i.e., resources and related consequences in OPTEMPO and demonstrated combat capability. An important by-product of this examination would be to improve our ability to identify indicators of current and future readiness.
3. Examine indicators for redundancy, i.e., identify those indicators that tend to vary consistently with each other and, thereby, add little additional information about status and trends. These indicators are candidates for elimination.
4. Examine indicators that could be combined by appropriate statistical procedures, perhaps increasing the reliability of the information and reducing the number of indicators to which senior decision makers must attend.
5. Examine the relation between subjective and objective indicators of readiness in an effort to identify the extent to which both are needed and whether the subjective assessments provide information not otherwise available.
6. Start the collection and analysis of new demonstrated performance measures such as percent of crews qualified, percent of Training Readiness Examinations above average, percent of Operational Readiness Inspections rated excellent or outstanding, and percent of mission essential tasks trained to standard.

I. INTRODUCTION

The purpose of this paper is to review training readiness indicators used now by the Services, Joint Staff, and the Office of the Secretary of Defense and to propose indicators that could improve current methods of reporting training readiness. The current method of reporting, called the Status of Resources and Training System, or SORTS, has been criticized because it relies on various subjective, rather than objective, assessments of training readiness and because it focuses on current readiness and does not provide estimates of training readiness in the near future.

Readiness is, as the Deputy Under Secretary of Defense for Readiness, describes it:

> the general ability of forces to arrive where they are needed, on time and prepared to effectively carry out assigned mission objectives for which they were designed. The ability of units to be ready on time to carry out their missions, in turn, is a function of having the equipment, supplies, logistics and experienced people with the skills to accomplish assigned tasks. (Finch 1996)

In this paper, we are concerned primarily with measures that estimate the skill and experience of military people and units to accomplish assigned missions, in short, indicators of training readiness. By law, the military Services, under the guidance and oversight of the Secretary of Defense, are responsible for training and providing combat-ready forces to the combatant Commanders-in-Chiefs (CINCs) who, in turn, are responsible for conducting the combat and noncombat missions assigned to them. Thus, the Services, the Secretary of Defense, the Chairman of the Joint Chiefs of Staff, and the regional CINCs need and rely upon regular reports of the training readiness of the forces.

These reports are used to assign or shift forces to various contingencies on the basis of their current readiness. The reports also show, obviously, where additional training is needed. The Services regularly provide many reports on their readiness with respect to training, equipment, logistics, and the like. Here, we are concerned with improving the quality of reports dealing primarily with training readiness, without detracting from the importance of other readiness reports.

That the conduct of war requires the efforts of all four military services is not a new idea. In 1946, General Eisenhower wrote to Admiral Nimitz:

> Separate ground, sea and air warfare is gone forever. If ever again we should be involved in war, we will fight it in all elements, with all services, as one single concentrated effort. (Eisenhower, 1946)

In "Joint Warfare of the U.S. Armed Forces," issued after the Gulf War, General Colin F. Powell, Chairman of the Joint Chiefs of Staff, said, "Joint warfare is team warfare" and "Joint Warfare is essential to victory."[1] In the current edition, General John M. Shalikashvili, adds "...all commanders must understand, teach, and apply joint doctrine as they prepare and train the men and women who wear America's uniform to fight our Nation's wars."[2]

This paper contributes specifically to the training component of the Joint Readiness Baseline Project, sponsored by DUSD (Readiness), to develop more objective measures of training readiness; the other main effort on this project concerns Resources for Training, a task being conducted by the Logistics Management Institute.

In the SORTS, the individual Services report their readiness to The Joint Staff. SORTS is generally acknowledged to be limited because it reports training accomplishments rather than demonstrated performance capability. Some of its reports rely heavily on commanders' subjective assessments of overall unit readiness. Training to perform Joint tasks is not explicitly tracked. A more detailed description of SORTS is given below.

To avoid later confusion, it may be helpful to point out that there are three types of training readiness indicators—input, process, and output—each with particular utility for some part of the training community. In essence, this follows a typical production model, with phases of input, process, and output:

Input: Includes resources required for training, such as fuel, repair parts, exercise ranges, depot-level repairables and consumables. Costs of these resources are largely found in the Operations and Maintenance budget and are of primary interest to the budget and planning communities and, ultimately, Congress. An important distinction should be made between costs allocated to training and, after the fact, how these funds were actually spent. Good examples of how O&M funds for training migrate to other uses may be found in the financing of contingency operations in Haiti and Bosnia.

Process: This concerns how much and what kinds of training actually took place (in contrast to budget allocations), in such terms as vehicle miles, flying hours, and steaming days used for training, called Operating Tempo (OPTEMPO). Such data reflect the extent to which specified training plans actually occurred and for

---

1 "Joint Warfare of the U.S. Armed Forces," Joint Pub 1, 11 November 1991.

2 Ibid, 10 January 1995.

which the term "training accomplishments" appears appropriate. Training accomplishment data are reported by training personnel and are of interest not only to the training department but to the commanding officer of each unit and to higher commands.

Output: Even if the amount of funds allocated and training accomplished imply a potential for achieving a particular level of training readiness, the practical issue is to know the level of combat performance demonstrated by units as a result of such training. Performance data may be subjective or objective, but objective data are strongly preferred. The primary type of subjective performance data is a commander's assessment, as part of his report to higher commands, of how well his unit performed in a particular exercise, or contingency, together with observations on any needed improvements and how to achieve them. Objective data, attainable primarily from exercises, include accuracy of gunnery and bombing, time needed to accomplish various types of maneuvers, and missions and force exchange ratios (kills and losses) in specified exercises. Table I-1[3] illustrates the types of information relevant to input, process and training output.

Table I-1. The Training Model

Input	Process	Output
Resources Costs	Training Courses Events Exercises OPTEMPO Accomplishments	Performance Measures of effectiveness Exercise outcomes Performance to standards Training readiness
Programmed/Actual Current and Future	Programmed/Actual Current and Future	Programmed/Actual Current and Future

All Services maintain extensive information, much of it available in automated databases, on their training accomplishments. Some training performance data, at Service and Joint levels, are available, but they are largely derived from commanders' assessments and are mainly subjective in nature; objective performance data, to the extent they exist, tend to be collected and maintained only for limited periods of time in the operating units. Thus, these data are difficult to access and to organize for analytical purposes.

Data on the costs of training are, most surprisingly, not readily available. The Logistics Management Institute summarizes its initial findings on training costs:

> Analysis of information readily available within DoD highlights the major difficulties in deciding how to best allocate the limited resources available for unit training. Our analysis focused on the operations and maintenance

---

3 Variations of this figure may be found in Prettol, et al. (1995), Jareb, et al. (1994) and Burba, et al. (1994).

(O&M) costs associated with unit training. (O&M costs include fuel, repair parts, depot-level repairables, and consumables.) We found that these costs for training are difficult to ascertain from existing data bases.

The programming process provides data in an aggregate form, but lacks specificity. Within the Future Years Defense Program (FYDP), operating costs are subsumed within program elements and resource identification codes. Operating tempo funds that include most training costs cannot be extracted from the larger O&M accounts. Furthermore, the FYDP contains no information about unit training requirements.

The budget displays also are of limited value in understanding the unit training process. The Services provide separate operating tempo budget displays that identify operating costs for major units such as ships and aircraft. Training costs, however, are not distinguishable from other operating costs, such as those associated with routine forward deployments and operations other-than-war.

High-quality accounting data reflecting training efforts are very limited. At the unit level, some training execution data are collected, but are held only for a short time and are not reported to others beyond the installation level. When available, some of the data may have value in explaining how training resources are executed, but considerable work would have to be done to clarify the definitions and interpretation of the data.

Our initial conclusion is that DoD has no systematic way of identifying the training requirement, the resources allocated to training, or the training accomplished. (Prettol, et al., 1995, p. iii–iv.)

Military training proceeds in stages: individual skill, crew, unit, service task force, and joint. All training except joint training is accomplished within each Service. The amounts and costs of residential individual skill training at various schools are reasonably well known and are reported annually in the Military Manpower Training Report; the amounts and costs of on-the-job training in units are not reported. It is estimated that 175,000 man-years of residential individual training will be provided in Fiscal Year 1997 at a cost of $13.7 billion (Department of Defense, 1996). All other types of training, called collective training, are conducted in operational units; these costs are not reported in any systematic fashion.

Joint exercises have been estimated to cost $0.4 billion in FY 1994; OPTEMPO has been variously estimated at $9 billion (FY 1993) or $21 billion (FY 1991), according to different analysts; and unit training has been estimated at $12 billion (FY 1993). It is not clear whether any of these sums can be added to yield a total cost of collective training (Orlansky, et al., 1994).

Additionally, funds appropriated for collective training may be diverted to other uses, adding further confusion to what is believed to be the total cost of collective training. The GAO found that in Fiscal Year 1993, the Army spent about $1.2 billion, or 33 percent of its OPTEMPO funds (part of the Operations and Maintenance budget), for other purposes, such as base operations, real property maintenance, and contingency operations in Somalia and Haiti (GAO, April 1995). No reductions in readiness were reported by the four divisions affected for seven of the eight quarters examined; in reporting degraded readiness during the fourth quarter of FY 1994, two cited a lack of funding.

Several initiatives at the Joint level and at least one at the Service level are intended to improve our current knowledge of the actual readiness of our forces at the Joint level. These include:

- Senior Readiness Oversight Council (SROC)
- Joint Monthly Readiness Review (JMRR)
- Joint Training System (JTS)
- Status of Readiness and Training System Improvements (SORTS)
- Joint Military Essential Task Lists (JMETLs)
- Joint Readiness Baseline Project
- Training Council for Modeling and Simulation
- Joint Simulation System (JSIMS)

The Commander-in-Chief, U.S. Atlantic Fleet (CINCLANTFLT), has initiated a pilot program under the Government Performance and Results Act (GPRA) to extend the SORTS format to reporting the readiness of ships in the George Washington Battle Group. A Carrier Battle Group consists of diverse surface, subsurface, and aviation assets, and if this pilot program is successful there is little reason to believe that it would not work for Joint forces as well. The system uses performance indicators, called B-ratings, based on such factors as equipment operability, Mission Capability Rates, operational availability, data link effectiveness, and measured air intercept performance of battle group units to rate the Battle Group's ability to carry out required mission tasks and subtasks.

II. STATUS OF RESOURCES AND TRAINING SYSTEM

This chapter describes the SORTS used by the Services and the Joint Staff to report training readiness and other components of readiness. It also considers proposals made to improve SORTS and analyses that examine the reliability of certain indicators to predict training readiness.

Service units report monthly on their overall combat readiness (C-rating) as well as on their readiness levels in four areas that contribute to overall combat readiness:

- Overall Level (C)
- Personnel (P)
- Equipment and supplies on hand (S)
- Equipment and condition (R)
- Training (T)

The current status in each area is reported as P, S, R, and T levels. Training levels T-1 to T-4 are defined by criteria shown in Table II-1; T-1 is ready for combat, while other levels are less ready.

Table II-1. Converting Days of Training or the Training Percentage into a T-Level

RULE	Resource Area Status Level			
	T-1	T-2	T-3	T-4
1. Days of training required	<=14 days	>14<=28 days	>28<=42 days	>42 days
2. Percentage of operationally ready aircrews for assigned personnel	>=85%	>=70%	>=55%	<55%
3. Percentage of mission-essential tasks trained for assigned personnel	>=85%	>=70%	>=55%	<55%

Note: Level T-5 is used to indicate that a ship or squadron is inactive because of scheduled major refurbishing or otherwise.

Source: CJCS, Joint Pub 1-03.3 (1993), p. XIV-32.

Note that Rule 2 refers to operationally ready aircrews and that Rule 3 refers to mission-essential tasks. This leaves open the possibility that the Services do not necessarily use the same rules to determine their T-levels. In addition, the commander of each unit provides an Overall or C rating, according to the following criteria:

a. C-1. The unit possess the required resources and is trained to undertake the full wartime mission(s) for which it is organized or designed.

b. C-2. The unit possesses the required resources and is trained to undertake most of the wartime mission(s) for which it is organized or designed.

c. C-3. The unit possesses the required resources and is trained to undertake many, but not all, portions of the wartime mission (s) for which it is organized or designed.

d. C-4. The unit requires additional resources or training to undertake its wartime mission(s), but it may be directed to undertake portions of its wartime mission (s) with resources on hand.

e. C-5. The unit is undergoing a Service-directed resources action and is not prepared, at this time, to undertake the wartime mission (s) for which it is organized or designed.

(CJCS Joint Pub 1-03.3, August 1993, pp. XIV-4 to XIV-5)

If a unit is not fully ready (C-2 or lower), the reason for that condition must be reported. Although commanders are directed to report the lowest P, S, R, or T ratings, the C-ratings are subject to change, based on the commander's Overall assessment.

The Gulf War crisis, August 1990 to February 1991, offered a limited opportunity to compare a unit's readiness for combat, as reported in SORTS, with a Service's willingness to commit that unit to combat. Specifically, the Army indicated that three National Guard brigades would require over 120 days of post-mobilization training even though their Commanders reported in SORTS that only the standard 40 days predeployment training would be needed to prepare their brigades for combat duty (GAO/NSIAD, September 1991; GAO/NSIAD, March 1992). The Army's assessment proved to be pessimistic because the three brigades were found to be ready in 90 days, 1 day before the war ended. Since then, a number of reports have identified various limitations in the DoD readiness reporting systems, with a primary focus on SORTS. This approach overlooks the fact that of over hundreds of units of all Services rated as ready in SORTS, only three were not deployed to the Persian Gulf. This suggests that SORTS is a more reliable

reporting system than its critics assert. The following criticisms have been made of SORTS:

1. The system, although mandated by CJCS, reflects unit (i.e., Service) rather than Joint readiness. Joint combat capability, observable in Joint exercises, is not reported in SORTS.
2. SORTS reports generic readiness, rather than CINC mission-specific readiness. Its structure and format do not use the recently adopted standard of Joint Military Essential Task Lists (JMETLs), used now by all regional CINCS for training in their assigned missions.
3. SORTS describes current readiness; it does not include estimates of future readiness over periods of, e.g., 6 months, 12 months, etc.
4. SORTS does not distinguish between conducting required training programs (i.e., process or training accomplishment) and demonstrated combat capability (i.e., output or performance effectiveness). The results of joint exercises are reported in the Joint Universal Lessons Learned System (JULLS) but not in SORTS. These results are narrative reports not designed for inclusion in a numerical data base.
5. Significant items in SORTS, such as Overall combat capability (i.e., the C-level ratings) can be based on commanders' subjective assessments, rather than on objective, demonstrated performance capability. Greater use of objective measures is now feasible and generally available in data compiled on instrumented ranges and in command post exercises that use combat models.
6. SORTS does not include the following information regarded as central to current and future readiness:
    - Mobility (Mobility is a Navy Primary Mission Area reflected in SORTS)
    - Morale
    - Leadership
    - Command, control, communications and intelligence
    - Exercises
    - Funding for training and OPTEMPO

This list might be extended, and many of the reported limitations of SORTS, particularly with regard to Joint capability, have obvious merit. However, few (or none) of those who would expand SORTS consider the relative importance, value, or cost of adding additional reports—or even the reliability of the data the reports would provide—especially in such areas as morale and leadership and command and control. In any case, except for

the three National Guard Brigades, other troops rated as ready for combat were actually sent to the Gulf and performed well.

The General Accounting Office (GAO, 1995, p. 3) reports that 28 active and reserve Service commands monitor over 650 readiness indicators in addition to SORTS. It reports that service officials said that 26 indicators (used by 2 to 16 of the 28 commands) are critical or important to predicting readiness (GAO, 1995, Attachment 1). These proposed indicators fall in the following categories:

category	number of indicators
personnel strength	4
personnel turbulence	4
personnel morale	1
training (including funding)	7
equipment fill	2
equipment condition	3
equipment maintenance	3
supply	2
Total	26

Similar suggestions are reported by Robinson, Jondrow, and Wheeler (1994), who interviewed 66 of the Navy's highest ranking officers, enlisted personnel, and civilian leaders. Although the focus was on how to prevent a hollow force rather than on indicators of readiness (or its potential lack or hollowness), the following elements were judged to be critical in avoiding hollowness.

Table II-2. Critical Elements in Preventing Hollowness[1]

Critical Area[2]	First Choice	In Top Three
Personnel quality	24	40
Total manning	8	19
Retention	3	18
Flying/steaming time	5	16
Training	2	21
Spares	2	15
O-level maintenance[3]	0	10

1 Taken from Robinson, Jondreau, and Wheeler, 1994, CRM 94-167

2 Other critical areas: Money (2), Force structure (2), Flexibility, and PERSTEMPO. Not all respondents chose to answer this question.

3 Operational-level maintenance

Even though these indicators appear, at least intuitively, likely to influence readiness, no analysis was provided to show that variations in any of them are consistently related to variations in readiness. Such work needs to be done before anyone suggests that adding any of these indicators would improve our ability to evaluate current or predict future readiness. In the area of training, with which this paper is most concerned, the GAO suggests seven new indicators:

Table II-3. Training Indicators Proposed by GAO[1]

10	Unit readiness and proficiency	Inspections, evaluations, and exercises including Combat Training Center rotations used to assess how well the unit is prepared to perform its mission
11	Operational tempo	Level of operational and training activity against specific standards
12	Weapon systems proficiency	Certifications, qualifications, and other indicators of individual and crew proficiency in military operations and weapons employment
13	Funding	Current and projected funding available for operations, training, and maintenance in units
14	Completion of required and specialty training	Numbers and/or percentages of personnel completing required or specialty training in a specific period
15	Commitments and deployments	Number and types of missions/commitments that (1) require all or part of a unit's resources or (2) do not provide an opportunity to train in all essential unit tasks
16	Accidents	Percentage of accidents in relation to standard measures, e.g., accidents per 100,000 flying hours

[1] Source: General Accounting Office, 1995

A. TRAINING RESOURCE MODELS

Models that attempt to relate training resources to current and future training readiness for combat are not new. In 1979, Horowitz and Hibbs (1979) reviewed 131 studies that examined some aspect of the resources-to-readiness paradigm, some dating back to the early 1960's. In 1978, the Army Training Study developed the Battalion Training Model (BTM), a computer model that related the amount of resources, in terms of time, dollars, people and facilities, required to accomplish what has to be trained. Readiness level would be measured as the additional number of training days required before deployment in a crisis. The highest state of readiness would be 0 additional training days needed; most units fell between 0 and 90–120 additional training days needed to be ready. The model was also capable of estimating the effect of varying levels of turnover on

training readiness. It could estimate trade-offs between (1) funds spent on recruiting or on retaining higher quality soldiers using pay and benefits and (2) paying civilians to perform selected tasks. The computer model appeared to produce intuitively accurate results. The estimated relationships between resources (as input) and readiness (as output) were based entirely on expert judgments and were not, in any way, tested empirically by comparing records and trends in funding against observed performance levels or judgments of unit readiness (as the Congressional Budget Office did in a report considered below).

Nevertheless, the Army leadership decided not to use or further develop BTM.

> First, it was felt the model would have insufficient credibility within the Army's field and higher level leadership because there are so many complex, even unponderable variables, in the training system that defied modeling (p. 3) despite the judgment of officers and civilians that the model was based on realistic relationships actually used by senior commanders. The second reason for discarding the model was:
>
> ...the fear that it could and would be used by analysts outside the Department of Defense for their own purposes (i.e., to find spurious ways to reduce the defense budget or make outrageous pronouncements about readiness status after the input of significant resources, etc. In the Army, no comprehensive effort to tie training resources and external influences [to readiness] has occurred since the ARTS model. (personal communication from a participant, now a retired senior flag officer)

Junor and Oi (1996), of the Center for Naval Analyses, examine analytically the extent to which various resources are reliable indicators of or influence the four areas of Personnel, Equipment, Supply, and Training Readiness. Overall Readiness, the commander's assessment of unit readiness, was not included in this analysis. For purpose of analysis, Junor and Oi define readiness as the percentage of time in a quarter that a unit is reported as C-1 in each of the four areas considered by SORTS. The historical database contains quarterly status reports on nearly every ship in the Navy, from 1978 through 1994. The method of analysis allowed for the possibility that the four resource areas are not independent. Table II-4 identifies the resources that were examined for each area of readiness. It is obvious that most readiness levels are influenced by more than one type of resource. All of the results are statistically significant at the 5-percent confidence level, except for personnel quality, which is significant at the 10-percent level. Some indicators that Junor and Oi regard as inputs—equipment failure rate, mean time to correct CASREPS (casualty reports), and percentage of time in C-1 for supply, and for equipment—are, we believe, outputs.

Table II-4. Inputs (i.e., indicators) That Significantly Influence Navy SORTS Readiness Levels, 1978–1994 (Based on Information Reported by Junor and Oi, 1996)

Inputs (Resources)	Outputs (SORTS readiness ratings[1])			
	Personnel	Equipment	Supply	Training
Personnel quality	X	X	X	X
Manning	X	X	X	X
Equipment failure rate		X	X	
Weapons procurement			X	
Ship spares		X	X	
Shore supply			X	X
Steaming	X	X		X
Deployed status	X	X	X	X
Deployment cycle		X		X
Overhaul cycle		X		
Modernization		X		X
Equipment age		X		
Equipment				X
Ammunition				X
Time	X	X		X
Crew turnover	X			X
Days under way this quarter	X		X	
Number of new C3/C4 CASREPS		X	X	
Repair parts			X	
Gross effectiveness			X	
Cost of scheduled maintenance last year			X	
Quarters since ships were last deployed			X	X
Approaching decommissioning		X	X	
Mean time to correct CASREPS		X		
Percentage of time in C1 for supply		X		X
Percent of time in C1 for equipment				X
Days underway last year				X

All Xs are significant at the 5-percent confidence interval, except for personnel quality, which is significant at the 10-percent level.

1 Time a ship spends in C1 for that resource area.

These findings suggest that SORTS readiness levels, even if the subjective assessments are judged not to be reliable, generally parallel trends in more objective measures of readiness. Personnel number and personnel quality strongly affect all areas of readiness. Indicators based on various types of resources can provide reliable indicators of current and future readiness, although future readiness is not explicitly considered in the Junor and Oi study.

This study is an important—and basically rare—contribution to our understanding of readiness indicators. Some areas that were not covered still require examination, such as

- Predictive value of various resources as to Overall Readiness levels, i.e., commander's assessment or C-ratings
- Time delay between use of particular resources (input) and their observed effects on readiness status (output). This is important for distinguishing between indicators of current and future readiness as well as for understanding how long it may take to improve various aspects of readiness
- Effect of increases in selected combinations of resources on readiness, to determine ways of making major improvements in readiness, if needed

This study was limited, by design and sponsorship, to the Navy. Similar efforts could be applied to the unit readiness reporting methods of other Services as well as, ultimately, to Joint readiness.

In 1994, the Defense Science Board Task Force on Readiness, composed of eight senior retired flag officers of all services and chaired by General Edward C. Meyer, USA (Ret.), found that

> ...the readiness of today's conventional and unconventional forces is acceptable in most measurable areas. That does not mean that the Task Force did not find 'pockets' of unreadiness. Most of these 'pockets' are a result of changes taking place in the armed forces and the turbulence created by these changes. (Memo, Meyer to Chm, DSB, 21 June 1994). There is, however, a need to support 'Development of measurement systems that better equate readiness to resources—present and future. The Department should take actions to develop and improve the set of analytical tools and other means that can be used to help better understand the relationship between funding allocation decisions and future force readiness' (DSB, p.11)...key indicators that measure readiness and provide early warning of potential readiness problems are strongest as they relate to a unit's current readiness within its Service and weakest as they address future and joint readiness. (Defense Science Board, 1994, p. iv)

The General Accounting Office (March 1996) reviewed data on and trends in military readiness for the period of January 1990 to March 1995 and reached a similar conclusion. It examined readiness reports for 94 units of all Services and found that most of the units remained generally stable and combat ready for assigned missions. In this study, the GAO used but did not assess the reliability of SORTS data provided by the Services.

In 1991, the GAO found that evaluations of the proficiency of Army units are not always reliable (GAO, February 1991). This finding was not the result of an evaluation of SORTS. Rather, GAO found that training at home station or even in exercises was conducted under conditions not sufficiently realistic to yield reliable information on a unit's combat capability.

B. STUDIES AND ANALYSES

In this section, we review studies and analyses that report statistically reliable indicators of training readiness. Few of them are oriented towards SORTS. All of them evaluate the predictive power of certain activities, such as flying hours, tank miles driven, or personnel turnover, on the level of performance of some military task, such as bombing, air-to-air combat, or torpedo attacks. Most findings are derived from records of training; a few are based on actual combat.

We have found 19 studies that show statistically significant relationships between various indicators of training and some objective measures of performance. Many of these studies examined flying hours and found that they predict performance on such objective, combat-related measures as bombing accuracy on instrumented ranges, quality of landing on carriers, and air-to-air combat exercises; two cases of bombing accuracy show the effect of training hours by Air Force and Navy pilots on combat missions in the Gulf War. For the Army, tank miles driven in training predict combat-exchange ratios in two-sided combat exercises at the National Training Center. Similar findings apply to Navy steaming days and SORTS C-1 ratings. Average length of time of officers in command of maneuver battalions was related to number of casualties suffered in combat by battalions in Vietnam (more time, fewer casualties). Table II-5 summarizes the findings of these studies; some indicators related to manning levels and crew stability are included because training readiness cannot be properly considered independently of personnel factors.

Most readiness indicators have not received such analytical treatment and their validity, even if we anticipate them to be significant, needs to be demonstrated. Among the

indicators that can be recommended for such examination, most relate to specific measurable variables whose contribution to success in battle remains to be demonstrated, e.g., successful landings on carriers, effective navigation, and Operational Readiness Evaluation scores. We also need a metric so that each of these different types and measures of performance can be transformed to a common scale that shows their contribution to overall readiness. As of now, all of these indicators refer to unit training readiness within the Services, and not yet to Joint readiness.

Table II-5. Some Readiness Indicators that Reliably Predict Level of Performance[1]

Indicator	Service	Effect on Performance (readiness status)	Source
flying hours and boarding rate on carrier	Navy	An increase of 80 flying hours per month per squadron (from 400 to 480 hours) raises average boarding rates on carriers from 90.1 to 93.2 percent.	Cavalluzzo, 1984
bombing accuracy	Navy	Average miss distance: a 1-percent increase in bombing flying hours is associated with a 0.5-percent decrease in average miss distance.	Cavalluzzo, 1984
Operational Readiness Evaluation score	Navy	Grades on Operational Readiness Evaluations: 20 percent fewer monthly pre-ORE flying hours per month (Pacific vs. Atlantic Fleets) associated with 39 percent of Pacific Fleet squadrons receiving scores in top two ORE categories, compared to 63 percent for the Atlantic Fleet.	Cavalluzzo, 1984
air-to-air combat	Navy	A drop of 10 percent in flying hours of F-14 fighters would increase the probability that red defeats blue by 9.2 percent and decrease the probability that blue kills red by 4.8 percent.	Hammon and Horowitz, 1990
landings on carriers	Navy	A 10-percent drop in flying hours would increase the number of unsatisfactory landings by F-14 and A-7 aircraft on carriers by 9.5 percent.	Hammon and Horowitz, 1990
bombing errors	Navy	As flying hours during the past week increased from 5 to over 20, bombing errors dropped from about 120 to 60 ft for AV-8B aircraft. Similar results for F/A-18 and F-4S.	Hammon and Horowitz, 1990
torpedo attack	Navy	A 10-percent decrease in pilot career flying hours decreases P-3 aircraft torpedo attack scores by 0.6 percent; a similar decrease in recent flying hours decreases these scores by 0.04 percent.	Hammon and Horowitz, 1996

(cont'd)

Table II-5. Some Readiness Indicators that Reliably Predict Level of Performance (Continued)

Indicator	Service	Effect on Performance (readiness status)	Source
flying hours and bombing miss distance	Marine Corps	A 10-percent cut in flying hours by AV-8, F/A-18, and F-4S aircraft would increase bombing miss distance by 1.8 percent.	Hammon and Horowitz, 1992
tactical air drops	Air Force	A 10-percent drop in career and recent flying hours by co-pilot and navigator of C-130 aircraft would increase average miss distance in tactical air drops by 2.8 percent.	Hammon and Horowitz, 1992
accident rates for aircraft	Air Force	Accident rates rose when flying hours were reduced during the late 1970's and dropped when flying hours rose in the 1980's.	cited in Hammon and Horowitz, 1990, p. 27
bombing accuracy	Air Force	Average bombing accuracy increases as career flying hours increase and levels off at about 900 hours for the F-16 and at about 1,400 hours for the A-10. Pilots above this threshold were better than those below them. Increases in squadron average flying hours per month from about 10 to 40 increase relative bombing effectiveness from about 0.35 to 0.95, linearly for the A-10 and asymptotically for the F-16.	Cedel and Fuchs, 1986
(last 30 days) and switchology errors	Navy	Reduces switchology errors. Increases number of excellent and satisfactory landings. Improves plan execution in air combat maneuvering. Increases skills in outer air battle. Launches HARM[2] correctly.	Newett, Davis et al., 1991
(last 30-90 days) and switchology errors	Navy	Reduces switchology errors. Improves number of excellent landings. Improves plan execution in combat air maneuvering.	Newett, Davis et al., 1991
(career hours) and switchology errors	Navy	Reduces switchology errors. Increases number of excellent and satisfactory landings. Improves plan execution in air combat maneuvering. Improves weapon employment score in war at sea.	Newett, Davis et al., 1991
air-to-ground bombing	Navy	Decreases bombing miss distance.	Newett, Davis et al., 1991
hitting targets in Gulf War	Air Force	On about 1,700 combat missions by F-117 aircraft during the Gulf War, every additional combat mission raises probability of a target hit by about 0.6 percent.	Gilman, Hammon, et al., 1997

(cont'd)

Table II-5. Some Readiness Indicators that Reliably Predict Level of Performance (Continued)

Indicator	Service	Effect on Performance (readiness status)	Source
flying hours and hitting targeted bridges in Gulf War	Navy	One-hour increase in flying time per month during 5 months prior to combat increases hit probability against bridges by 15 percent for unguided bombs; this effect is not observed for laser guided bombs probably because of larger "drop basket" and corrections by guidance system.	Gilman, Hammon, et al., 1997
average length of time of officers in command of maneuver battalions and battle deaths in Vietnam	Army	Maneuver battalions in Vietnam (34 battalions, 1965-1966) under experienced commanders (6 months or more in command) suffered 2/3 as many battle deaths as battalions with commanders with less than 6 mos. in command.	Thayer, cited in Tillson and Canby, 1992, p. III-4
number of flights in combat and probability of being shot down	United States and Korea	Probability of being shot down in combat decreases from about 0.40 to 0.25 as number of combat flights increases from 1 to 10.	H.K. Weiss, 1966
air combat losses	Army Air Force	During World War II, the air combat loss rate dropped from about 20 percent on the first combat flight to about 2 percent on the 16th combat flight.	H.K. Weiss, 1966
number of flights in combat and air combat losses	German Air Force	During World War II, the air combat loss dropped from about 40 percent on the first combat flight to about 2 percent on the 21st combat flight.	H.K. Weiss, 1966
Red Flag training and Top Gun training for war in Vietnam and kill exchange ratios	Air Force Navy	As a result of air combat training on instrumented ranges, Navy and Air Force kill ratios in combat over Vietnam rose from 2.4 in 1965-1968 to 12.50 in 1970–1973.	P. Gorman, 1990
number of undersea combats, World War II and submarine kills	Navy	After one kill, a submarine commander increases his chances of further success, as opposed to chances of losing his submarine, by a factor of 3.	H.K. Weiss, 1966
manning level of ships and serious failures	Navy	An increase of 34 percent (one standard deviation) in enlisted ship manning relative to requirements increases percent of time a ship is free of serious failures by 6 percent.	Beland and Quester, 1991
crew stability on ships and serious failures	Navy	An increase of 34 percent (one standard deviation) in enlisted crew new to the ship this quarter reduces the percent of time a ship is free of serious failures by 4 percent.	Beland and Quester, 1991

(cont'd)

Table II-5. Some Readiness Indicators that Reliably Predict Level of Performance (Continued)

Indicator	Service	Effect on Performance (readiness status)	Source
miles driven during 6 months prior to exercises at the NTC and force exchange ratio	Army	More miles driven improves performance at NTC in force-on-force offensive mission (correlation: r = 0.68) and in live-fire defensive missions (correlation r = 0.80). Most successful brigade drove about 780 miles in training, least successful about 380 miles.	Holz, O'Mara and Keesling, 1994
days underway per quarter and aircraft navigation	Navy	Scores on low-visibility piloting exercise: an increase of 1 day underway (from 30 to 31 days) raises score by 0.2 points.	Follman, Marcus and Cavaluzzo, 1986
fuel budget and C-1 rating	Navy	Fraction of time in C-1 on SORTS: An increase of 1 percent in the fuel budget is associated with an 8-percent increase in C-1 time and a decrease in all lower readiness categories.	Cavalluzzo, 1984
personnel turnover and C-1 rating	Navy	Probability of deploying C-1 for training: For ships in first deployment since overhaul, a decrease of 1 percentage point in new crew (from 11.8 percent) increases probability of deploying in C-1 for training by 2 percentage points (from 82 to 84).	Marcus, 1989

[1] Incorporates and supplements information presented in Robinson, et al. (1986), Appendix B. Quantified Estimates of Effect.

[2] High-speed Anti-Radiation Missile.

[3] National Training Center.

The sections that follow contain summaries of relevant studies on the effects of training on readiness. These studies were conducted by the Center for Naval Analyses (CNA), the Congressional Budget Office (CBO), the U.S. Marine Corps (USMC), the U.S. Army (USA), and the U.S. Navy (USN).

1. CNA Studies

A series of studies at the Center for Naval Analyses examined various relations between the budget for Naval training and readiness. Training accomplishment reported in UNITREP (a predecessor system similar to SORTS) was found to be associated positively with the fuel budget, i.e., the level of OPTEMPO predicts ship readiness (Cavalluzzo and Morben, 1982; Cavalluzzo, 1985). Demonstrated performance would be a more valid indicator of training readiness than training accomplishment, whether measured by fuel used or number of steaming days. For example, using a sample of 134 ships over the

period of 1982 to 1985, OPTEMPO was found not to be associated with whether or not a ship passed its mobility inspection (Follman and Cavalluzzo, 1986). This strongly suggests—but does not prove—that some training effectiveness measures are not just poor discriminators of performance, as these results show, but may be inherently unreliable indicators of what they purport to measure. The same study showed that over a period of four Atlantic fleet competitive cycles (1979–1985), the probability of winning mission-area awards increased with OPTEMPO until reaching a peak at about 40 steaming days per quarter, after which point it declined. In this case, the criteria included both training accomplishment and exercise performance. Although OPTEMPO improves both (up to a point), its contributions to process and output were not separated in the analysis.

Cavalluzzo and Horowitz (1987) developed and tested multivariable models to examine hypothesized linkages between resources and readiness. The study uses data on Navy ships from UNITREP and other sources during the period of 1977 to 1985. Various segments of the study examine relations between data on costs, training accomplishments, demonstrated performance, and C-ratings. The paper makes the following points:

- Overall status of resources (C-ratings) can be used as the single leading indicator of training accomplishment; it is influenced strongly by equipment condition and personnel resource levels because adequate training cannot occur when such deficiencies are present.
- The level of appropriations for readiness implicit in the budget accounts for operations and support are positively associated with training accomplishments.
- Training accomplishments increase with OPTEMPO and real expenditures for fuel.
- The association between OPTEMPO and demonstrated performance remains to be demonstrated.
- Univariate analyses that attempt to relate OPTEMPO to performance have had mixed success, but much of the blame may lie with the quality of the data used to describe performance.
- The study used what is believed to be high quality performance data—scores recorded by independent observers at gunnery ranges. In this analysis, equipment design, rather than personnel and training levels, proved key to performance.
- The strong association between OPTEMPO and training accomplishment combined with the weak association between OPTEMPO and performance

remains to be reconciled in future research. (Cavalluzzo and Horowitz, 1987, p. ES-9, 10)

Newett et al. (1991) examined the relationship between flying hours of Navy F/A-18 pilots and their performance in selected mission areas. Data were collected on two fighter/attack squadrons over a period of 1 year. Three measures of flight hours were used:

- recent flying hours (within the last 30 days)
- semi-recent flying hours (in the last 30 to 90 days)
- career flying hours

Performance data were collected in three mission areas:

general	number of switchology errors number of communication errors overall flight procedures navigation procedures carrier landing grades
air-to-air	escort/air combat maneuvering outer air battle
air-to-ground	high-speed anti-radiation missile opposed bombing unopposed bombing war at sea

Statistically significant relationships ($p < 0.10$) were found between recent and career flying hours and various measures of performance in each mission area; these are included in Table II-5. Many of the measures follow a classical learning curve: performance improves as flying hours increase up to some level and then flattens out. Although all pilots improve at about the same rate, as flying hours increase, appreciable differences in level of performance remain between the best and worst pilots, regardless of number of flight hours.

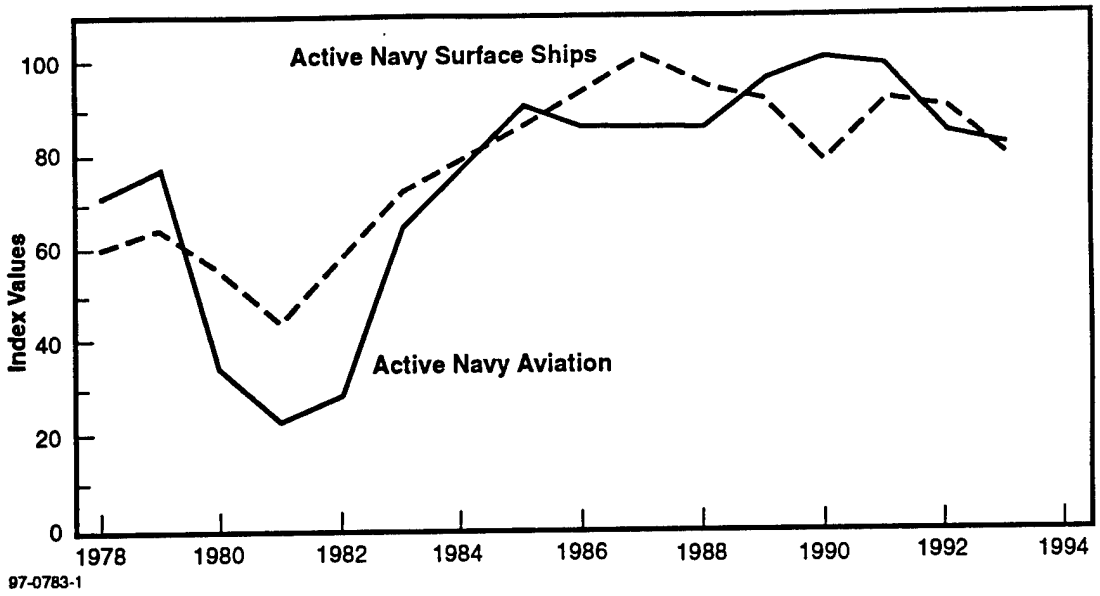
2. CBO Studies

The Congressional Budget Office (CBO) examined trends in selected indicators of military readiness during the years of 1980 to 1993 (Congressional Budget Office, 1994). The period of 1979 to 1982 is taken to represent a case of low readiness, i.e., the "hollow force," while the period of 1990 to 1992 represents a case of high readiness, particularly as demonstrated by our success in the war against Iraq. The key question that CBO addressed was the extent to which trends in various SORTS indicators reflect the reality of these two levels of our military readiness. This paper, together with others cited in this chapter,

offers an opportunity to validate or at least to improve our understanding of some readiness indicators.

Figure II-1 shows that C-ratings for Navy aviation and surface ships were low and high, respectively, during the two critical time periods; similar trends are found for the other Services. However, the condition of Army and Marine Corps ground equipment was fairly stable over the same time period (Figure II-2). Therefore, some limitations in these output indicators may be noted. Some output indicators did not rise as fast as increases in relevant resources might have suggested because different time intervals are needed to produce observable results in such areas as training, purchases of spare parts, and depot level maintenance. Some apparent changes in readiness are actually due to changes in reporting rules. For example, in the early 1980's, the Navy encouraged commanders to report, rather than downplay, problems in readiness. One paradox is that the introduction of new and advanced equipment may, at first, actually reduce readiness until crews are fully trained to handle the new equipment and the necessary supply and maintenance support systems become effective. In 1986, Secretary of Defense Weinberger asked the Joint Chiefs of Staff to change the terminology used to describe C-ratings in a way that emphasized its dependence on resource levels and de-emphasized its potential as an indicator of military readiness.

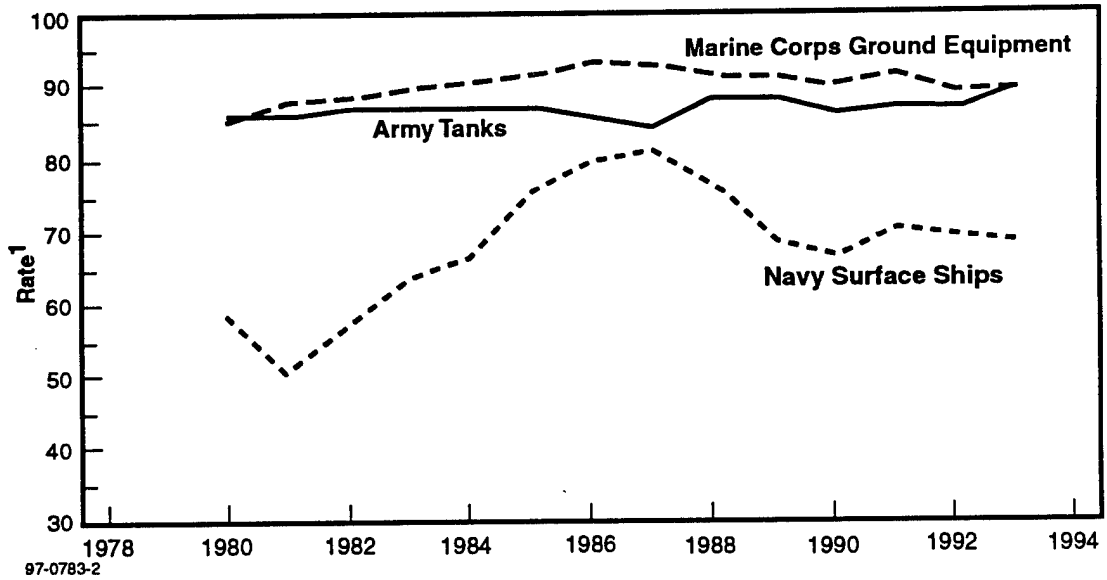
The CBO also reviewed trends in resource indicators that might be linked to future readiness levels, such as the ability to recruit and retain high quality personnel, changes in experience level of personnel, the availability of sufficient personnel, and the match between available personnel and needed skills. Beyond the area of personnel and training, similar problems exist in trends for spending on operations and maintenance, depot maintenance backlogs, and the availability of supply. The CBO found limitations in such potential indicators of future readiness. For example, retention rates among career military personnel are driven by DoD policy and not by the influence of morale on retention or the attractiveness of military compensation. This does not mean, of course, that retention rates or morale cannot be useful indicators of readiness. Highly aggregated indicators of unit readiness may be unreliable, compared to more specific and detailed ones. CBO suggests that DoD focus on developing better indicators of current unit readiness before attempting to identify early-warning indicators. It emphasizes the need for more objective performance measures, the use of performance evaluators from outside the unit being evaluated, surprise or limited-notice evaluations, and the use of measures that are directed or weighted towards the unit's assigned combat missions.



Source: Congressional Budget Office (1994) based on Department of Defense data.

Note: These indexes show changes in the percentage of units reporting C-1 or C-2 relative to their peak values; they do not show the actual percentage of units that are C-1 or C-2. The peak value for aviation units was in 1990, so the aviation index in that year has a value of 100. The peak value for surface ships was in 1987, so the surface ship index has a value of 100 in that year. An index value of 50 for a given year means that the percentage of units reporting C-1 or C-2 was half of its peak value.

Figure II-1. Indexes of C-Ratings for Navy Units, 1978–1993



Source: Congressional Budget Office (1994) based on Department of Defense data.

1. Army and Marine Corps: Mission capable rate Active component only.
   Navy: Percentage of operating time free of C-3 or C-4 equipment casualty reports (CASREPs) Active component only.

Figure II-2. Condition of Ships and Ground Equipment, 1980–1993

3. USMC Studies

Jareb, et al. (1994) examined the following tasks for the Marine Corps:

- Relate the unit training readiness of Marine Corps ground battalions to training resources.
- Identify better measures or indicators of unit training readiness.
- Examine how changes in training resources affect unit training readiness.

The study effort was limited to the training readiness of infantry, artillery, and tank battalions. According to the authors, unit readiness is determined by several elements that include, but are not limited to, personnel readiness, equipment, availability of ranges and facilities, and time available for training.

Limitations in the current SORTS readiness reporting can be addressed by adding indicators of:

- Performance to unit training standards (individual training standards already exist)
- Personnel stability (average length of time in this unit)
- Unit status in the training cycle
- Subjective assessment of the commander in areas difficult to evaluate by objective measures

This paper provides estimates of the costs of Marine Corps operations and maintenance expenditures (not all for training) in units exercises, ammunition, training ranges, other base support, and equipment maintenance. Ammunition accounts for 43 percent of all costs, followed by maintenance (28 percent). As welcome as these approximations are, the authors observe that

> The quality and detail of the data we received on the costs of unit training vary considerably. In many cases, Marine Corps cost accounting procedures and systems do not permit precise or accurate estimates of unit training costs. Also commands and units have some latitude in how they group expenses for accounting purposes. For example, some units may track fuel costs in a separate account, whereas others may include fuel costs as one of several expenses in a single account (Jareb, et al., 1994, p. 61).

In the authors' judgment, some of these resources have effects that are observable only over the long run:

- Expenses for equipment maintenance,
- Expenses for training ranges and facilities, and
- Other base support activities

Some expenses—unit O&M funds, exercise funds, and ammunition—are expected to produce more immediate results. Although intuitively attractive, no evidence is offered to support any of these hypotheses. The authors are undoubtedly correct in their judgment that no single measure can adequately assess unit training readiness.

4. USA Studies

Hiller, McFann and Lehowicz (1994) examined the relationship between OPTEMPO at home station and performance of brigades at the National Training Center. Data were available for 16 brigades on task mileage (proxy for amount of training) during the 6 months prior to engaging in exercises at the National Training Center and on the casualty exchange ratios (OPFOR/BLUFOR) they experienced during the exercises. About 10,000 MILES rounds were fired in 58 defensive and 42 offensive exercises. MILES, the Multiple Integrated Laser Engagement System, uses lasers on guns and laser detectors on tanks to assess the accuracy of direct-fire weapons. There was a statistically significant correlation of 0.68 ($p = \leq .002$) between tank miles while training and casualty exchange ratios in exercises. Thus, for Army brigades tested in combat scenarios at the NTC, more OPTEMPO predicts better combat performance.

For several reasons, this is a remarkable finding. The primary finding, of course, is that using objective performance data, there is such a large, positive, and significant correlation: prior tank mileage acounts for almost 50 percent of the variance related to performance at the NTC. Still, tank mileage, per se, tells us nothing about the types and amounts of training associated with large (or small) expenditures of fuel. It appears that battalion commanders who used more fuel had good reasons—associated with training—for doing so. The "Determinants" study, described next, confirms these findings.

In the "Determinants" study (Holz, O'Mara, and Keesling, 1994), the approach was to compare measures taken on units at their home stations before visiting the NTC, with performance measured against the OPFOR at the NTC. The NTC is an instrumented range that can record extensive objective information on maneuvers, firing accuracy, and casualties in highly realistic, two-sided engagements between visiting brigades (the BLUFOR) and the resident, highly experienced opposing "enemy" force (OPFOR). The basic problem is to estimate how well training influences performance in live combat exercises. The collection of similar data in actual battle, however desirable, is rarely accomplished for both obvious and not obvious reasons (see Horowitz, et al., 1995).

Data on seven brigades included vehicle miles driven (OPTEMPO) during the 6-month train-up period and casualty exchange ratios (performance effectiveness measure) during offensive and defensive exercises at the NTC. The correlation between OPTEMPO and performance for offensive missions was 0.68; for live fire defensive missions it was 0.80. Earlier, Hiller, et al. (1994) found correlations of 0.0 for offensive missions and 0.69 for defensive missions. The most successful of the seven brigades had driven an average of 750 miles per vehicle in the 6 months prior to training, while the least successful brigade drove 380 miles. Hiller, et al. also show that there is a relationship ($r = 0.56$) between the similarity of the terrain on which units train and that of the NTC; home station areas are always smaller than that of the NTC and therefore much less favorable for offensive scenarios.

There is no doubt that factors such as leadership, morale, and group cohesion strongly influence the performance of troops in combat and in developing their training readiness for combat. Leonard Wainstein (1986) cites numerous cases from World Wars I and II where troops suffered large-scale casualties but on account of superb leadership and morale continued to fight effectively, and won battles against larger enemy forces. In a similar view, there is no doubt that the success of the 2nd Amored Cavalry Regiment in the Battle of 73 Easting during the Gulf War was due to conscientious training and superb leadership at all echelons (Orlansky and Thorpe, 1992). It is difficult, but not impossible, to "measure," (i.e., estimate reliably), such factors as leadership, morale, and cohesion and to incorporate them as reliable subjective assessments in a readiness reporting system. It is also reasonable and feasible to use some objective proxy measures of these factors in such variables as average rotation of personnel, average time of officers and enlisted personnel in a unit, and percent of qualified crews in a unit. For example, Thayer, in an analysis of the Vietnam War, 1965–1972, reports that "maneuver battalions under experienced commanders (6 months or more in command) suffered battle deaths in sizable skirmishes at only two-thirds the rate of units under battalion commanders with less than 6 months in command."[4] Stability in battalion commanders (those with more than 6 months in command) was associated with 32 percent fewer battle casualties per month compared to battalions with leaders who had less time in command.

---

4 Thayer, Thomas C., editor, *A Systems Analyst's View of the Vietnam War, 1965-1972*, Volume 8. "Casualties and Losses," February 1975, p. 225 (DTIC AD A051613), cited in Tillson and Canby (1992), page III-4.

The Army has established some stabilized tank platoons where there is less rotation of personnel, and units, rather than individuals, are replaced or rotated. Tillson and Canby (1992, p. III-5) report that 75 percent of stabilized platoons achieved the highest scores in UCOFT exercises (a tank gunnery simulator) while only 15 percent of the nonstabilized platoons met this standard.

These empirical findings strongly suggest that one or more indicators of stability of personnel (e.g., long tenure within units) would be reliable predictors of training readiness (and probably performance in actual combat). The probability of success in the Gulf War was probably enhanced by eliminating the routine rotation of personnel during the period of Desert Shield.

III. TRAINING READINESS DATABASES

A. INTRODUCTION

This chapter describes existing service databases which contain information on training readiness. The databases report training accomplishment and performance at the unit level and above which are or can be summarized at the major command level. The databases also provide building blocks for a system to track trends in training readiness. Current training readiness indicators provide a snapshot in time of existing unit readiness levels, but there is no systematic means of capturing trends or of predicting future readiness. To meet these goals, we have selected a small number of training readiness indicators that draw on existing information which, to the extent possible, report training accomplishment and performance rather than subjective assessments.

1. Where Are We Now?

The only source of systematic information on training readiness available to the Office of the Secretary of Defense is SORTS. SORTS C-ratings are generally based on accomplishment of required training and are easily understood. However, SORTS has several important deficiencies:

- SORTS describes current readiness; it does not estimate future readiness.
- SORTS describes unit readiness; it does not describe Joint readiness.
- SORTS data are particularly subjective for ground forces and rely heavily on the Commander's assessment of overall unit readiness.
- SORTS provides little indication of important differences among individual units and little indication of important trends.
- SORTS reflects inputs to training as well as outputs, but does not systematically identify particular input deficiencies.

The Services have various fine-grained indicators which can be used at the major command staff level for assessing the training readiness of their units. For the purpose of reporting to the Office of the Secretary of Defense, we have attempted to identify

aggregated databases that provide data on unit performance, but which do not identify individual units.

2. Criteria for Assessing Training Readiness Indicators

The following criteria were used in assessing training readiness indicators:

- They should be easy to understand, particularly at upper levels of the Services, OSD, and the Joint Staff. This, of course, does not mean that they lack substance and real informational content.
- They should not add to the reporting burden of individual units.
- There is no need to identify individual units in reports to OSD and the Joint Staff.
- The data should reflect training accomplishment and performance in separate categories.
- Training readiness indicators should be objectively determined. They should be measured against a standard rather than reflect the relative standing of like units. In addition to objective data, human judgment is useful in the determination of whether training accomplishment or performance is up to standard. There should be well-defined and well-known performance standards and criteria for comparison.
- Indicators should be discriminating and consistent. To be useful as indicators of training readiness, databases should discriminate among various levels of readiness and do so consistently. One of the frequently cited problems with SORTS is that as training budgets and physical resources change, readiness reports may not follow the change or even move in the same direction.
- Training readiness indicators should provide broad coverage across services and kinds of units; the sampling frequency should provide frequent snapshots of a representative part of the force. This does not mean that the actual kinds of data must be the same for all services and units. It does mean that the indicators used for each service and unit must provide consistent data streams.

Archiving such data provides the basis for observing trends, establishing limits beyond which various changes can be noted as significant, and ensuring that we have the information needed to relate changes in resources to changes in readiness.

3. Categories of Training Readiness Information

After consideration of resources, training readiness indicators fall into two basic categories: training accomplishment and training performance.

As described earlier, training accomplishments refers to the types and amounts of training that are planned or conducted; this is the process phase of a production model. Training performance refers to the level of performance demonstrated or assessed as the result of training; this is the output or product phase of a production model.

The amount of resources allocated to various types of training may predict near-term and future readiness, on the assumption that specific relationships of this type can be demonstrated, but detailed training resources are not considered in this chapter.

4. Training Performance Measurement

Training accomplishment data document how much training has been performed. This category includes OPTEMPO measures and specific training events completed. OPTEMPO measures include flying hours, steaming days, vehicle miles, training ordnance expended, and range hours used. To be useful, of course, these must be measured against some standard. OPTEMPO measures are useful in conjunction with other data which specify how the time and resources were used. Such data may be specifically recorded, they may be based on averages, or they may be related to the accomplishment of training events. Specific training events completed include school attendance, squadron attendance at Top Gun or Strike University, battalion attendance at the National Training Center or Twenty-nine Palms Combined Arms Exercise, etc.

Using objective measures or using carefully defined subjective assessments by trained observers, training performance is measured by how well required tasks or missions are conducted. Although the observer assessments are a subjective category, their reliability can be established through well-defined rating procedures and training of observers to score such events. Subjective assessments are potentially valuable, provided that explicit reference material has been developed and pre-tested for commanders to use as guides for what is to be reported as excellent, adequate, or inadequate performance.

Objective training performance reports show how well a task is performed. This category implies quantitative and objective measurement. These reports include data such as the results of air combat or bombing exercises conducted on instrumented ranges, Table VIII or Table XII firing scores, NTC force exchange ratios, scores for gun or missile shoots, etc. Unfortunately, objective training performance data are often limited to only part of a unit's total mission. For example, bombing errors accurately measure one part of the air-to-ground mission, but do not indicate proficiency in finding, entering, and egressing a hostile target area.

5. Approach

First, we identified variables which are or can serve as measures or substitutes for training readiness. We tried to identify measures of training accomplishment and performance used to track training readiness and available at the major command levels of each service. We accomplished this primarily through interviews with service training and operations personnel at several command levels, ranging from operating units through type commanders and major commands. We also reviewed service training directives and interviewed members of the Joint Staff.

In our interviews and literature searches we tried to identify databases which were consistent with the criteria stated above, with particular emphasis on performance data. Although we initially looked for a wide range of information maintained at many levels, our final list focused on a small number of promising sources used at the type command and major command level for tracking training readiness.

The questions addressed included:

- What is the nature of the data? Where and in what format does the data originate? What is the initial purpose of collecting the information and who eventually uses it? Does the information refer to process or output?
- How are the data gathered and how reliable are the source documents?
- Are the data based on independent evaluations?
- Where are the data located?
- Is the database automated? Who has access to it?
- What summary measures of training readiness can be developed from the data?

The following section describes databases of each service that meet these criteria.

B. SERVICE TRAINING READINESS INFORMATION

This section addresses the small number of databases available at the major command level that, in accordance with criteria established above, have merit as objective indicators of training readiness.

1. Sources of Army Training Readiness Information

The Army includes standardized evaluations as part of their overall training system. However, the results of evaluations are closely held and there is a reluctance to release information, particularly quantitative information on performance, to higher authority or

OSD. In addition, many exercise reports avoid grades (perhaps purposely) in favor of narrative comments and lessons learned. The consensus seems to be that if units know they are being assigned numerical or descriptive grades, or that the quantitative evaluations will go beyond the immediate superior, there will be an adverse effect on the unit and its commander. Personnel will somehow attempt to game the system to produce favorable (rather than actual) ratings, rather than concentrate on learning from training deficiencies.

Sources of training readiness information include the following:

- External evaluations in the Army Readiness Training Evaluation Program (ARTEP)
- Crew qualification levels
- Participation in exercises at Combat Training Centers

a. Army Readiness Training Evaluation Program (ARTEP)

ARTEP is the basis for essentially all Army unit training. ARTEP provides a list of mission elements and subelements (mission essential tasks) that each kind of unit (crew/squad through battalion) should be able to accomplish to carry out its mission. Mission essential tasks are now being related to the JMETLs of the Joint Staff and CINCs. ARTEP evaluations are conducted at all echelons for both active and reserve components. External evaluations are conducted and observed by the chain of command every 12–18 months.

External ARTEP evaluation data are categorized as training performance evaluations. Units are evaluated as "Trained," "Partially trained," or "Untrained" on each mission essential task. Observers are supposed to evaluate each unit relative to a standard rather than to other like units. However, as with many performance evaluations, there is no effective mechanism to ensure that this is done. Nevertheless, external ARTEP evaluations are probably the most useful and objective training readiness indicator available at the battalion level. Battalion records are maintained at the division level. Currently, the results are being maintained in the Standard Army Training System (SATS) automated database, but coverage is only partial.

A potential training readiness indicator would be the percent of mission essential tasks evaluated as "Trained" within and across units. Units need not be identified by name. This measure could be aggregated to yield the percentage of like units scoring a certain percentage of mission essential tasks as "Trained." Since ARTEP evaluations are a snapshot in time and are repeated only every 12–18 months, the measure would be a

sample of the total force; of these, some ratings would be recent, some old—but on average, they report combat capability 6–9 months ago. Even so, a report of each quarter's activity would indicate trends in the training readiness of the force.

b. Crew Qualification Levels

Tank and Bradley crews conduct sustainment and attainment exercises on a continuing basis. Sustainment exercises represent day-to-day training. Attainment exercises are graded and successful completion represents a qualification. A successful firing on Table VIII, for example, represents crew qualification; platoons qualify on Table XII. The results of these exercises are categorized as objective training performance data, the most desirable type of indicator. The percentage of crews qualified is reported quarterly to the division commander. OPTEMPO information (e.g., vehicle miles, ordnance expended, etc.) is also reported quarterly.

Crew level qualification training includes many of the exercises conducted in support of the Battalion Level Training Model (BLTM) program. The BLTM, which uses OPTEMPO and Standards in Training Commission (STRAC) standards, is intended to relate training activities to the level of training readiness based on the number of trials required to maintain various levels of training readiness.

Potential training readiness indicators include percent of crews qualified on Table VIII and percent of platoons qualified on Table XII. A possible weakness of these measures is that they do not report the number of attempts needed to qualify. The system is intended to train units during day-to-day sustainment exercises and to conduct attainment exercises when the units are trained and can be expected to qualify on the first attempt. Supplementing the measures with STRAC and OPTEMPO data would be useful in this regard.

c. Participation at Combat Training Centers (CTC)

The Army considers participation in combined exercises at one of the Army CTCs or the National Training Center the best training the Army can provide. Armored and mechanized battalion task forces (Brigades) train in live-fire and force-on-force engagements against a resident and well-trained opposing force (OPFOR). Exercises involve the combined operations of tanks; mechanized infantry; artillery; air defense; engineers; electronic warfare; simulated nuclear, biological and chemical warfare; attack helicopters; and fixed wing aircraft in close air support. Both active and reserve components participate

in CTC exercises, which are conducted without breaks and without outside support. Units must therefore demonstrate full logistics capability, including maintenance, evacuation of casualties, and field billeting and messing.

A potential training readiness OPTEMPO indicator would be the CTC personnel-days in a quarter, as a fraction of total personnel-days in the force. This measure is classified as training accomplishment. Although the measure is highly aggregated, in a budget climate characterized by limited resources it might provide early warning of decreasing training opportunities and, conceivably, decreasing actual readiness.

CTC exercise data are obtained by monitoring radio transmissions and by using fixed and mobile video cameras. Digital data are obtained from the MILES. Soldiers and vehicles are fitted with sensors that register hits and near misses by laser beams which simulate direct fire of weapons. Assessment of damage by noninstrumented weapons, such as mines and grenades, are made by observers. Data are recorded in the field using electronic clipboards—small laptop computers which provide menu-driven information—into which data can be fed using electronic pencils. The devices, which are carried by each observer/controller, check lists and rules of engagement nearly instantaneously.

The NTC Analysis Division of the Center for Army Lessons Learned (CALL), located at Fort Leavenworth, is responsible for collection and analysis of CTC data. However, the data are primarily subjective, and reports do not contain any reference to actual events or units. The observers and CTC staff conduct debriefings and prepare a comprehensive take-home package intended to assist the exercise units in correcting deficiencies after they leave the center. The take-home packages are maintained only by the participating units.

The CTCs, while potentially the most objective source of performance data for large scale ground warfare exercises, have not been used to build a comprehensive database. The reasons for this appear to be both technical and doctrinal. On the technical side, the MILES coverage does not include all elements of the exercise: it does not always penetrate smoke and dust generated on the battlefield and it cannot reach units masked by terrain. In addition, the Observer/Controllers may affect the outcome of the exercise by their efforts to stimulate interaction with the OPFOR and enhance training. On the doctrinal side, the Army has stated that the primary mission of the CTCs is training, and its position is that emphasis on the collection of objective data would be detrimental to that mission.

There is at least one known exception to this rule. The Army conducted a study that showed that the number of vehicle miles driven by brigades during the 6 months prior to going to the National Training Center correlated highly with their performance in offensive and defensive force-on-force missions at the NTC (Holz, O'Hara, and Keesling, 1994).

Exchange ratios are one example of objective data that could be collected during CTC rotations. Unfortunately, the exercise rules, training objectives, and environmental variables are not constant for each exercise. Still, averages across rotations might be meaningful as a training readiness measure. Average exchange ratios at the National Training Center and other CTCs are a potential training readiness indicator categorized as performance measurement. Their use would probably meet with heavy resistance because of concern over interference with the training mission cited above and the absence of sufficient controls, such as standardization of exercise conditions.

2. Sources of Navy Training Readiness Information

All Navy tactical training—surface ships, aviation units, submarines, and tactical staffs—is based on a Tactical Training Strategy designed jointly by CINCPACFLT (Commander-in-Chief Pacific Fleet) and CINCLANTFLT and promulgated in the Tactical Training Manual.[5] The training strategy supports naval warfighting capabilities at the Unified CINC level and below. Required naval capabilities, as stated in the JMETLs, are translated into training policy at the Fleet CINC and numbered Fleet level. Specific guidance is given to deploying units through operational tasks (OPTASKs) based on their Required Operational Capabilities (ROC) and Projected Operational Environment (POE). ROC and POE state the required operational capabilities of units by warfare area.

The Navy training and readiness system is based on an 18-month deployment training cycle. Toward the end of each deployment, the Commanding Officer conducts a training assessment to determine the amount of schooling and other training asset requirements. Following a month of reduced operations at the end of each deployment, interdeployment training is divided into three phases—Basic, Intermediate, and Advanced—ending with a Final Evaluation Period immediately before deployment to the Sixth or Seventh Fleet. Required exercises and certifications of increasing complexity and difficulty are conducted throughout these three phases. Each phase of the training cycle includes a performance evaluation. Training readiness data are maintained by the Type Commander

---

5 CINCPACFLT/CINCLANFLT INSTRUCTION 3501.1.

(TYCOM) throughout the interdeployment training cycle; the TYCOM also allocates resources for training.

Sources of Training Readiness information include:

- TRMS, Type Commander Readiness Management System (surface including aviation ships)
- Navy Aviation Training and Readiness Matrix
- Training Readiness Exam (subsurface)

a. Type Commander Readiness Management System (TRMS)

TRMS is a PC-based system which will eventually be accessible to all ships as well as Type Commanders and CINCLANT/PACFLT. TRMS is based on the Tactical Training Manual reporting requirements. When a ship completes a required exercise or certification it sends a Training Report message to the TYCOM which includes scores, who observed, etc. From these messages the TYCOM upgrades the unit's mission-specific readiness rating (M-rating) and when the exercise needs to be repeated, if necessary. M-ratings are measures of the unit training readiness by Primary Mission Area (PMA) and are aggregated to compute the unit's training status in SORTS. The TYCOM or fleet CINC can then query the database to see what each unit's training status is in terms of percentage trained by PMA. The system consists of several Modules:

- Executive Information System (EIS)—This module is now operating. It is intended to give Commanders a quick summary of force readiness and training status.
- Casualty Report (CASREPT) Module—This is primarily, but not exclusively, used by the logistics people. This module also provides summary information used by COMNAVSURFLANT to track the operational status of all ship propulsion, machinery, electronic, and weapon systems of ships throughout the interdeployment training cycle and while deployed. This module is operational.
- Training Readiness System Module—This module is up and running but has not been distributed to all units. Its main purpose is to assist units in submitting Training Report messages and to facilitate review of their status. When completed, each ship will be able to access the central database to confirm its M-ratings and training status. At the present time, ships receive feedback reports once a month.
- Personnel Module—This module is currently being introduced at the TYCOM and Fleet CINC levels. The Personnel Module adds a great deal of detail and

is fully Windows compatible. The module uses data which was previously available at the Bureau of Personnel and at the Total Force Management Systems Command. It will be used to keep track of Personnel quantity and quality levels throughout the fleet.

- Inspection Module—This module enables the TYCOM and Fleet CINC staffs to keep track of the inspection status of each unit. Staff officers can look at a summary of inspection results and requirements, as well as grades and deficiencies of each unit. The module is now operating.
- Combat Systems Module—This will keep track of missile shoots, torpedo shoots, etc.
- Ships Readiness System—This module is close to being implemented. It is really the payoff module and ties the whole system together. The staff can call up summaries of the training and readiness status of units by whatever organizational breakdown they desire. For example, if they need a Carrier Battle Group (CVBG) for a special operation, this module is intended to let them review all CVBGs and select the one best suited.

TRMS data are categorized as training accomplishment and objective and subjective performance evaluations. Coverage includes continuously updated information for nearly all surface ships, including aviation ships, and it is the only database examined which meets all of the criteria listed above (paragraph III.A.2). Potential training readiness indicators include average percent trained by ship type and mission area, and average percent of certifications completed. The system is discriminating and consistent, and could support measurement of readiness of CVBGs and Afloat Training Groups (ATGs) for Unified CINC missions.

b. Naval Aviation Training and Readiness Matrix

The Navy manages and tracks air crew training through the ROC measure with percentage for each Primary Naval Warfare Mission Area (PMA). Training matrices for each aircraft type list training events (ROCs), the percentage applied to each PMA, ordnance requirements, and the interval for repeating that event. These, along with ground training and school requirements, form the framework of the squadron flight training plan and are used to measure and report training readiness. To evaluate air crew training progress, certain individual designations are also monitored, such as flight leader, special weapons delivery pilot, mission commander, and aircraft commander (for multi-piloted aircraft). In addition, Navy air crews must qualify in certain exercises to qualify in each PMA. When these exercises are graded by the type wing staff (or an observer designated

by the wing commander), they are called competitive exercises (COMPEXs). Completed COMPEXs are a potential training readiness measure.

A PMA is a primary mission area capability which a unit must have in order to carry out assigned tasks. ROC application percentage measures the percentage contribution of a specific training event to the capability of a unit to perform a given PMA. For each PMA the total of all ROC application percentages adds to 100. A flight crew member is considered to be combat ready in a particular PMA if his ROC application sub-percentage is equal to at least 75 points out of 100. Squadron M-ratings and percent of training accomplished by PMA are based on the number of qualified crews relative to the number of crews allowed. Squadron training progress is available in a PC database at the wing and type commander level. These data are updated through the monthly Training and Readiness Report.

The average percentage of training accomplished successfully, based on average percentage of qualified crews, flight hours, range hours, etc., is a potential training readiness indicator. This measure is categorized as training accomplishment. However, it includes certain performance measures, such as average circular error probable (CEP), submarine contact time, successful air-to-air engagements, successful COMPEXs, etc. which are used to calculate crew qualification.

c. Navy Submarine Training Readiness Exam

Submarine training readiness is reported by the squadron commander based on the results of a Training Readiness Exam (TRE) given every 12 to 18 months. This is unique to the Submarine force. All other ships and aircraft squadrons report their own SORTS status. The TRE is administered by a special team composed of members from the TYCOM and Squadron staffs. The TRE is given during a 3-day Composite Training Unit Exercise (COMPTUEX) which reflects the employment anticipated during the ship's next deployment. All departments and all aspects of the submarine mission are examined and adjective grades assigned for:

- Tactical mission
- Strategic mission
- Navigation
- Material
- Basic submarining
- Overall

Assigned grade categories for each of these areas are Excellent, Above Average, Average, Below Average, and Unsatisfactory. Portions of the exam, including torpedo and missile certifications, are conducted on an instrumented range. Numerical grades (Probability of kill) for these portions of the exercise are calculated by range personnel.

The data are located at the TYCOM headquarters in an automated database, where they are analyzed for trends. Potential training readiness indicators are percent of overall grades in each of the five categories and percent below average. The data are categorized as performance measures and contain objective performance measurement elements, such as torpedo and missile kill probabilities. The standards of performance are well defined and are generally very demanding. Samples of the data which we have examined follow a bell shaped curve and appear to be discriminating. The TRE is often, but not always, given in conjunction with a Predeployment Certification (PDC). The submarine force is unique among naval forces in that failure on the PDC usually results in the ship being pulled out of the normal rotation and another ship deployed in its place.

3. Sources of Air Force Training Readiness Information

The Air Force training and inspection system is similar to the Navy aviation training, but with more emphasis on inspections. Air Force units are expected to maintain level readiness, and major readiness inspection exercises are administered on a time schedule, rather than adeployment schedule. Inspections are the responsibility of the Inspector General, rather than the commands that do the training. Operational training programs are monitored at the Major Command (MAJCOM) level by the continuation training section of the Deputy for Operations (DO) as part of the Realistic Training Review Process. Sources of training readiness information are:

- MAJCOM Training Programs
- Operational Readiness Inspections (ORIs)

a. USAF MAJCOM Training Programs

Flight training requirements in the Air Force are very similar to the Navy Training Readiness Matrix. The MAJCOM develops unit Designed Operational Capabilities (DOC) statements based on the JCS Joint Strategic Capabilities Plan (JSCP). The MAJCOM continuing training division is responsible for designing training events to ensure that units are capable of accomplishing their wartime mission. The Wing and flying squadron

commanders are responsible for administering and monitoring the training program. Requirements are stated in terms of successfully completed sorties and events.

Air Combat Command (ACC) crews progress through a required series of events in order to attain progressively higher Graduated Combat Capability (GCC) levels. Succeeding levels (GCC level A through GCC level C) require proficiency in progressively more advanced events. Although all services adjust flight training hours according to the experience level of personnel, the Air Force is the only one that formally states different flight hour requirements for experienced and inexperienced air crew personnel.

GCC levels are defined as:

- Level A—Minimum qualified and mission ready in primary capabilities (equivalent to C-1 in SORTS)
- Level B—Level A plus trained to support specific tasks (the ACC goal is to have 80 percent of combat air crews at GCC level B or above)
- Level C—Qualified tasking for full capabilities

The wing keeps track of air crew progression through GCC levels as well as range performance data derived from gun camera film and electronics pods used on instrumented Air Combat Maneuvering (ACM) ranges. GCC level data are forwarded to the ACC/DO Realistic Training Section where they are kept in an automated database. The Realistic Training Section uses the data, along with monthly visits to operating squadrons, to monitor training readiness and to prepare inputs to the Realistic Training Review Board. Composed of representatives from Plans, IG, Safety, Intelligence, Guard, and Reserve, this Board meets every 6 months to act on issues submitted by the ACC/DO. The board has cognizance over:

- Revised mission statements
- Training event tasking messages
- Risk assessment actions
- Inspection scenarios

Potential training readiness indicators are:

- Percent of air crews qualified at GCC levels B or C (ACC)
- Percent of training program completed (ACC and AMC)

These are categorized as training accomplishment measures.

b. Inspector General Operational Readiness Inspections

The Inspector General (IG) inspection system includes both administrative inspections, called Quality Air Force Assessments (QAFA), and Operational Readiness Inspections (ORI). The QAFA addresses training accomplished, including percent of crews at GCC levels A, B, and C, numbers of flight instructors, flight lead, etc., and maintenance training accomplished.

MAJCOM/IG teams administer Operational Readiness Inspections every 2 to 4 years to active component wings. The goal is 2 years, but the time interval has been increasing with budget decreases. Regional IG teams administer ORIs for Guard and Reserve wings and NATO inspects USAFE (U.S. Air Force Europe) wings. The inspections are given in two phases.

Phase I: No notice for active component units.

Phase II: With notice; covers employment, mission support and ability to survive and operate.

Grades: Outstanding, Excellent, Satisfactory, Marginal, and Unsatisfactory.

Inspection results are maintained by the MAJCOM in an automated database.

Potential training readiness indicators are:

- Percent of evaluations in each category Air Force wide
- Percent Outstanding, Excellent, or Unsatisfactory

These measures are categorized as performance measures. A weakness of the system is the long time between inspections. Even if a return is made at a 2-year frequency, quarterly data would represent a small sample.

4. Sources of Marine Corps Training Readiness Information

The Marine Corps training readiness system contains elements similar to those for Army ground forces and Navy aviation systems.

All elements of the force, both ground and aviation, are evaluated by the Marine Corps Combat Readiness Evaluation System (MCCRES), similar to the Army ARTEP. Battalion-size units and above are evaluated externally, and the results are forwarded to the Marine Corps Combat Development Center (MCCDC). Tank crews and platoons qualify on Table VIII and Table XII, similar to the Army. Combined arms exercises are conducted at Twenty-nine Palms, but these are somewhat different from the Army Combat Training

Center exercises. The Marines conduct live fire exercises rather than opposed exercises against a resident OPFOR.

The aviation training and readiness system is very similar to that of the Navy, but with some refinements. The Marines have a system of squadron instructors similar to the Air Force. They also emphasize combined and air wing exercises earlier in the training cycle and more frequently than the Navy.

Sources of Marine Corps training readiness data are:

- External MCCRES evaluations
- Tank crew qualification levels
- Combined Arms Exercises (CAX)
- Aviation Training and Readiness Information Management System (ATRIMS)

a. External Marine Corps Combat Readiness Evaluation System (MCCRES) Evaluations

The mission of every ground and aviation unit is subdivided into tasks, and tasks are subdivided into requirements. Each requirement is defined so that its satisfactory execution can be described as "Yes" or "No." The contribution of each task and requirement to the overall mission is weighted in determining a score. MCCRES measures the performance of each unit under simulated combat conditions against well-defined standards, called Mission Performance Standards (MPS). There are 17 MPSs for an infantry battalion, divided into 4 sections: all evolutions, amphibious assault and normal combat operations, specialized combat operations, and use of outside support assets.

Unit MCCRES evaluations are conducted every 12–18 months and are evaluated by the next higher command. The results of these external evaluations are forwarded to MCCDC where they are maintained in a PC database. According to the MCCDC Training and Evaluation (MCCDC/T&E) division, the database is incomplete and 100 percent of all units perform to "standard." It is also conjectured that the lack of discrimination may be partly because the external evaluations are administered by the next higher level in the chain of command. This tends to be akin to self-evaluation. An effort is underway to correct the deficiencies of the MCCRES in particular and training support in general in a program called the USMC Training Readiness Support System (MCTRSS).

A potential indicator of training readiness is percent of mission essential tasks trained, broken down by like units. This is categorized as performance evaluation.

b. Tank Crew Qualifications

Tank crew and platoon qualifications on Table VIII and Table XII, respectively, are maintained in a PC database at the battalion level and forwarded to Division G-3. Qualifications are conducted in the same manner as for the Army, and the data have the same deficiencies as for the Army. The data do not address the number of attempts prior to a unit successfully qualifying. The system is intended to train units during day-to-day sustainment exercises and to conduct attainment exercises when the units are trained and can be expected to qualify on the first attempt. A current effort is under way to design a training and readiness system similar to the Marine Corps aviation system, ATRIMS, but the program has not been funded.

Potential training readiness indicators include percent of crews qualified on Table VIII and percent of platoons qualified on Table XII. These measures are categorized as performance measurement.

c. USMC Combined Arms Exercises (CAX)

Infantry and artillery battalions participate in combined arms live fire exercises at Twenty-nine Palms; these last 3 weeks for infantry and 6 weeks for artillery. MCCRES evaluations are not conducted during these periods, nor is there a dedicated OPFOR. Post-exercise reports are maintained at Division G-3 and are forwarded to MCCDC. These are written reports which are not retained in an automated database. There is concern at MCCDC about rotation frequency of exercises in the post-cold-war period. The goal is to have each battalion participate in a CAX rotation every 2 years. Over the past 2 years only 20 of 31 infantry battalions have participated—a rate of once every 3 years.

Frequency of rotation—percent of units or personnel participating in CAX during a given time period—is a potential training readiness measure. CAX participation is categorized as training accomplishment.

d. Aviation Training and Readiness Information Management System (ATRIMS)

The aviation training and readiness system is similar to the Navy's. In fact, for common aircraft such as the F/A-18, the Replacement Air Wing syllabi are identical. Marine and Navy Replacement Pilots train in the same squadrons.

Marine Corps squadrons train to some operational requirement by progressing crew members through a series of sorties, each of which incorporates required events (e.g.,

section takeoffs and landings, auto or dive bomb, 2 vs. 2 similar or dissimilar air-to-air, etc.). The training manual specifies the number and types of ordnance to be used for each event and whether an instrumented range is required. Required schoolhouse and other training efforts are specified, and each crewman must progress through certain stages—basic flight, wing man, section lead, flight lead, aircraft commander for multi-place aircraft, etc.

The Marines use a measure called Combat Readiness Percentage (CRP) to manage and track air crew flight training. The CRP for a given air crew reflects the percentage of syllabus flights in which the crew is current. Currency requires satisfactory initial completion and additional flights at specified intervals. Syllabus flights are categorized by the training stage in which they are flown. These stages progress from Combat Ready (60 percent–70 percent CRP) through Combat Qualification (71 percent–85 percent CRP) and Full Combat Qualification (86 percent–100 percent CRP). SORTS readiness levels are defined by the percentage of squadron crews which have attained a given CRP.

Tactical fixed-wing and helicopter squadrons participate in squadron, MAG, and combined arms exercises. For fixed-wing aircraft these exercises are the Fleet Fighter Air Combat Readiness Program, Strike Fighter Advanced Readiness Program training exercises, and Top Gun. For both rotary- and fixed-wing aircraft the exercises are Fallon and Yuma detachments and combined arms exercises. These exercises form an intrinsic part of interdeployment training and are therefore incorporated into the CRP ratings.

A potential readiness measure is the average CRP by aircraft type for pilots and Flight Officers in operational units. Average CRP is categorized as a training accomplishment, but because of the stringent well-defined standards and screening process in aviation squadrons, it incorporates elements of performance measurement.

C. SUMMARY AND FINDINGS

The training readiness of nearly all of the DoD combat force structure can be tracked using a few available indicators. These indicators, or derivatives of them, are for the most part available and used at major command levels to keep track of unit training progress. They are not always used in an organized way to track force training readiness in the aggregate.

Table III-1 is an overview of the databases examined in this report, categorized according to certain characteristics. The form and content of each database is dependent on the training and evaluation philosophy of each service, and varies across services. There is also considerable variation in objectivity and philosophy concerning the primary purpose of performance evaluations across the Navy surface and subsurface, aviation, and ground forces communities.

We found no system that uses current training readiness information as a baseline to predict future trends. The only system that shows real promise in terms of reporting training readiness according to tasks is the Navy TRMS. We have reviewed the capabilities of this system, but we have not had an opportunity to examine the actual data that TRMS appears to provide. According to a description of its capabilities, it can provide training readiness information applicable to SORTS at many organizational levels, down to exact training exercises (U.S. Navy, no date). It can identify areas of degradation and access to background information, such as personnel, ammunition, or equipment conditions, that needs remediation. A more complete description and analysis of TRMS data could be the subject of future work.

1. Ground Forces Training Readiness Indicators

The Army and Marines use similar performance evaluation systems for ground forces, including mechanized and artillery units. These systems vary in objectivity and in intent as to their proper application across types of units.

The Army ARTEP and Marine MCCRES evaluations are very similar in format and application. Both are used primarily as an integral part of the Service ground force training program. Both Services emphasize that the evaluations are intended to provide feedback to the evaluated units rather than training readiness indicators for higher command. The Army is particularly careful not to identify units or quantitative results to higher command. For both Services, external evaluations are conducted by the immediate superior in command. Although the percentage of mission essential tasks trained, as determined by external evaluations, is potentially the most objective measure of training readiness at the battalion level, this measure seems to be lacking in its ability to discriminate among like units.

Table III-1. Potential Service Training Readiness Indicators

Type of indicator	Indicator	Service	Units or personnel reported	Completeness of coverage	Location	Objective data?
Demonstrated Performance[1]	percent of crews or platoons qualified	Army	tanks and fighting vehicles	complete	division	yes
		Marine Corps	tank	complete	battalion	yes
	percent TRE grades above or below average	Navy	Submarines	complete		yes
	percent of ORIs excellent or outstanding	Air Force	all	sample	MAJCOM	yes
Training Accomplishment[2]	percent of mission essential tasks trained	Army	battalions	sample	division	apparently
		Marine Corps	all	sample	MCCDC	uncertain
	percent training accomplished by primary mission areas	Navy	surface ships	complete	FYCOM	yes
	percent of training accomplished (% Crews Combat Ready)	Navy	aviation Squadrons	complete	TYCOM	yes
		USMC	aviation Squadrons	complete	Wing	yes
	percent GCC level B or A	Air Force	aviators in ACC	complete	MAJCOM	yes
	percent participation in CTCs/CAX	Army	all	complete	Fort Leavenworth	yes
		Marine Corps	all	sample	MCCDC	yes

[1] Output: Demonstrated performance, e.g., percent targets hit, performance to standard, force exchange ratios in combat exercises.

[2] Process: Training accomplished, e.g., flying hours, vehicle miles, ammunition expended, percent air crews combat ready.

Abbreviations:

ACC	Air Combat Command
CAX	Combined Arms ExerciseAir Combat Command
CTC	Combat Training CenterOperational Readiness Inspection
GCC	Graduated Combat Capability
MAJCOM	Major Command
MCCDC	Marine Corps Combat Development Center
ORI	Operational Readiness Inspection
TRE	Training Readiness Examination
TYCOM	Type Commander

Table VIII and Table XII platoon gunnery scores are objective crew performance measures. They may not always indicate the level of training accomplishment. The system is intended to train units during day-to-day sustainment exercises and to conduct attainment exercises when the units are trained and can be expected to qualify on the first attempt. However, percent of crews qualified does not indicate the number of attempts required to qualify. A current effort is under way in the Marine Corps to design a tanker training and readiness system similar to the aviation system. Such a system would indicate actual training accomplished. This program has not been fully funded.

The percent by unit type of personnel or units participating in Army CTC or Marine CAX rotations over time is perhaps the best aggregated measure of training accomplished. Both services appear to have well-developed standards for these exercises. Performance measurement indicators exist for Army CTC rotations, but their usefulness is limited because of variations in exercise conditions and contraints on the availability of the data.

2. Navy Surface Training Readiness Indicators

The Navy TRMS is the most comprehensive and best suited database for the purpose of this paper. Indeed, the system was developed for the purpose of providing the Fleet and Type Commanders with the same information we are seeking for the Under Secretary of Defense (Personnel and Readiness). Using the various modules of the TRMS software, several options for training readiness indicators are available, including performance evaluations and measurements. In general, the most useful indicators translate into training accomplished by type of unit and PMA. Different operational task and organization levels can also be tracked. This is the only database we have examined that has the potential for tracking the training readiness of joint forces. Although the system is used primarily for surface ships, the basic design of the system is intended to accommodate all elements of a strike or amphibious Task Group, including submarines and ground and aviation units.

3. Navy Subsurface Training Readiness Indicator

The submarine force Training Readiness Exam database is perhaps the most discriminating of all the training readiness sources examined. It is also the only database which is currently being analyzed for trends in training readiness over time. However, these analyses are only just beginning and represent a small sample in terms of completeness of coverage over short periods of time.

4. Aviation Training and Readiness Matrices

Air crew training measures and tracking systems are similar for Air Force, Marine Corps, and Navy Aviation units. Navy and Marine Corps data may be retrievable at the type command or major command level, and perhaps at the headquarters level as well. All three Services train to some operational requirement by progressing crew members through a series of sorties, each of which incorporates required events. Each Service specifies the number and types of ordnance to be used and whether an instrumented range is required. All three specify required schoolhouse and other training evolutions and require that crewmen progress through certain stages—basic flight phase, wing man, section lead, flight lead, etc.

Potential measures of training readiness for all three Services are the percentage, by aircraft type, of aircrews that have attained a given level of readiness or the percent of training accomplished (GCC level). An additional indicator for the Air Force is percent of wings scoring excellent on ORIs.

5. Findings

The Services could be asked to report training readiness using a small number of Service-wide indicators. The indicators reported in this paper, or a subset of them, would serve as a starting point

The focus should be placed on percent trained to standard in a form suitable for each service, for example, average percent of required training accomplished for surface ships and aviation units broken down by PMA. It is true that percent trained does not directly address the question of how much training is enough. However, this question must eventually be answered based on the capability required of the force and human judgments as to how to attain that capability.

The strengths of the measures presented in this chapter are that they:

- Are objective
- Are understandable
- Incorporate evaluations and performance measurement
- Cover most of the force
- Require very little added reporting burden

These measures also have certain weaknesses. The measures are conceptually similar to SORTS. In their present form, the indicators and databases do not address the

dynamics of readiness. In most cases they represent snapshots in time and may require further analysis and some modification to be useful in a system which will show expected trends in training readiness. In narrowing the indicators to a small number, valuable additional sources of information are ignored. This could be mitigated by supplementing the list with additional information. The Services continue to develop these and other databases. In some cases, additional data collected and used at the unit level could easily be included in higher level databases. For example, the Navy/USMC flight information reporting system (NAVFLIRS) has a provision for recording ordnance expended and bombing accuracy, which are available when working on an instrumented range. This information is rarely reported but could easily be.

SORTS has been criticized because it contains subjective judgments. This is, of course, correct, because in the overall assessment of a unit's readiness level, i.e., the C-rating, a commander may report that his unit is combat ready (C-1) even though the unit is not fully ready in some area, such as personnel training or supplies; in such cases, the commander must report the reason for the lower condition(s). That the commander's rating is subjective does not necessarily mean that the judgment is inaccurate or unreliable, as implied by the criticism. The extent to which commanders' subjective assessments are unreliable can be determined empirically by a fact-finding experiment in which such judgments are compared to those of no-notice, independent operational readiness inspection teams. The critics have not, of course, done this.

Subjective assessments are not necessarily unreliable or inaccurate but easily become so if little attention is given to how they are collected and scored. Often, subjective judgments may be the only type of information that is available, either for lack of instrumentation for objective measurement or where the variable to be evaluated is not well defined or understood (such as, e.g., leadership or group morale). Reliable subjective measurement is possible provided that a scale has been developed that describes events that raters can observe when they score a unit on a scale of, e.g., 1 to 7 or Excellent, Good, Average, or Poor.

A valid criticism of SORTS is that it reports current readiness and does not attempt to estimate future readiness. This is an area where a commander's estimate of the future readiness of his unit 3, 6, or 12 months in the future may provide a useful early-warning indicator. On a test basis only, a commander could be asked to provide an estimate of the future readiness of his unit and the usefulness and accuracy of this information could be determined after a trial period of 12 to 18 months.

IV. DISCUSSION

SORTS is the principal report available to the Office of the Secretary of Defense and The Joint Staff on the training readiness of the four services. SORTS has been judged—not always fairly—by the General Accounting Office and the Congressional Budget Office as being based on inaccurate data and subjective assessments. As pointed out earlier, indicators of training readiness are of two types: those that can be described as "training accomplishments" and those that can be described as "demonstrated training performance." Accomplishments refer to the amount of training undertaken, measured objectively by such indicators as courses attended, number of flying hours or ship steaming days, ammunition expended, number of tasks trained, and participation in exercises. There are also some indirect indicators of training events, such as amount of fuel used and amount of ammunition expended. None of these measures define the level of performance achieved by the expenditure of, e.g., fuel, flying hours, or ammunition. Demonstrated training performance refers to how well specific tasks or missions relevant to combat are performed as measured by, for example, percent of mission-related tasks performed according to a formal standard; percent of shots or bombs hitting a target; or force-exchange ratios in two-sided exercises, whether live, virtual or constructive.

Note that a third type of indicator, called "resources allocated," is missing from this structure and is not considered in this paper. This category would include funds allocated for (1) flying hours or steaming days, (2) ammunition for training exercises, (3) training at schools, and (4) transportation of equipment and personnel to participate in combat exercises away from the home station. Because funds allocated to training accounts may be diverted to other purposes, analyses of the relation between resources and readiness should be concerned with actual, rather than allocated, expenditures for training. Funds allocated to training may be identified in the budget requests submitted annually to Congress and are centrally available. Funds actually spent for training may be found in records kept by the units being trained and in reports submitted to the Major Commands of each Service.

We found that the Services now collect certain information that is objective in nature and that are potential indicators of training readiness. These indicators are generally available at Major Command levels, and their use at Service Headquarters or in the

Department of Defense and the Joint Staff would not impose new data collection efforts. Based on the review described in Chapter III, we propose a set of indicators of training readiness. Table IV-1 lists these indicators along with the Services that now employ them; some can apply to all Services.

Table IV-1. Overview of Potential Training Readiness Indicators

Type of indicator	Indicator	Service
demonstrated training performance	percent of crews or platoons qualified	Army Marine Corps
	percent TREs above / below average	Navy subsurface
	percent of ORIs excellent or outstanding	Air Force
	percent of tasks trained to standard	Army Marine Corps
training accomplishment	percent of mission essential tasks trained	Army Marine Corps
	percent training accomplished by primary mission area	Navy
	percent of training accomplished (percent crews combat ready)	Navy aviation USMC aviation
	percent GCC level B or A	Air Force
	percent participation in CTCs/CAX	Army Marine Corps

CAX Combined Arms Exercise
CTC Combat Training Center
GCC Graduated Combat Capability
ORI Operational Readiness Inspection
TRE Training Readiness Examination

We know, also, that various studies show that certain indicators of training readiness, like flying hours, vehicle miles driven, and ship days underway, are statistically significant predictors of particular combat capabilities, as demonstrated by various objective performance measures. This list follows:

Flying hours	Navy, Marine Corps, Air Force
Vehicle miles	Army
Ship days underway	Navy
Personnel turnover	Navy
Length of time of officers in command	Army

We have shown that some of these indicators, or ones closely related to them, are statistically significant determinants of performance in exercises and, in a few instances, in combat. The validity of length of time of officers in command of a unit, as an indicator of training readiness, is based on combat casualties during the war in Vietnam; the validity of pilot flying hours as an indicator of training readiness is based on bombing accuracy during the Gulf War. There is no reason to believe that the validity of indicators based on data from only one Service would not also apply to similar activities in other Services.

All of these indicators reflect Service training readiness and not Joint training readiness. One may believe that the source of information needed for validation of these (or other) indicators for Joint training readiness lies in data collected in Joint and CINC exercises and real-world contingencies; this points to Joint after action reviews and the Joint Universal Lessons Learned System sources of information needed to verify the significance of some indicators. It may also be the case that data collection procedures for Joint and CINC exercises need to be developed and tested before reliable Joint level indicators can be developed.

It is also possible—in fact, probably necessary—to look at the development of training readiness indicators as a technical matter of measurement. The issue concerns the reliability and validity of measurement tools. The Services are not novices on this topic, as demonstrated by their successful development of instrumented ranges at the operational level and of sophisticated test facilities for evaluating developments in, e.g., ordnance, sensors, and engines. It would be helpful to propose a large variety of indicators that are considered relevant to readiness. Then, we face the nontrivial question of assessing the reliability and validity of each potential indicator. Only those that survive routine assessment for repeatability and relevance (i.e., validity) and are therefore demonstrably capable of serving as measurement tools need be considered further for acceptance or worth their cost. If DoD and the Services have a genuine need for improved readiness indicators, it will be necessary to support the technical efforts required to develop and prove their worth. Since the issue concerns assurance of readiness status, this hardly seems an arguable point.

V. RECOMMENDATIONS

SORTS has been criticized because it includes unreliable and subjective information on training readiness. Nevertheless, we found 29 analyses that show that certain indicators of training are valid (i.e., statistically significant) predictors of combat capabilities based on data collected in field exercises or in actual combat. These indicators are process-based measures, such as flying hours, steaming days, personnel turnover, and tank miles driven, that reliably predict objective combat capability, e.g., bombing accuracy in exercises and in war, success in air-to-air combat in exercises and in war, battle deaths in war, and C-1 ratings in SORTS. Though limited in number and not necessarily an endorsement of the entire SORTS system, these studies clearly show that certain SORTS indicators have demonstrable validity.

We must concur with suggestions to improve SORTS that call for more objective performance measures, the use of independent performance evaluators, i.e., from outside the unit being evaluated, surprise evaluations (a procedure practiced by General Curtis Le May and Admiral Noel Gaylor), and combat mission-oriented performance measures. The intuitively appropriate appearance of these suggestions is not, however, a substitute for more formal analytical evaluation of their utility, i.e., add only information which is worth its cost.

Some readiness indicators, particularly those associated with overall assessments made by senior officers, are subjective estimates for which no objective measures exist and which may not be feasible to develop. No suggestion is made here to eliminate or alter such assessments being made currently by commanding officers. These assessments may, in fact, be enhanced by adding features that could increase their reliability. The reliability and consistency of subjective measures can be improved by providing descriptions of what is meant, e.g., by "excellent," "good," or "needs improvement," so that all observers mean about the same thing when reporting a particular condition or status. Guidelines to subjective assessments could also suggest a list of items on which ratings are requested. Steps could be taken to extend current subjective assessments with a somewhat more organized reporting procedure.

Our examination of the training readiness reporting systems of the Services found candidate indicators that are objective, understandable and require little additional burden to be reported:

Demonstrated Performance

percent of crews qualified
percent of Training Readiness Examinations above/below average
percent of Operational Readiness Inspections rated excellent or outstanding
percent of mission essential tasks trained to standard

Training Accomplishment

percent of training accomplished by primary mission area
percent of training accomplished
participation in combat training center exercises
percent achieving Graduated Combat Capability level

These indicators have been used by the Services. Data, to the extent that it has been retained by the Services, should be available to assess the reliability and validity of these indicators as part of the process needed to recommend them for adoption.

We have examined the systems used by the Services to compile and report training readiness information. Of these, the Navy's TRMS has the most comprehensive database and its software is best suited for examining the utility of training readiness indicators. It appears to have the functionality of a system that could track the training readiness of Joint forces. It contains modules that provide data on equipment casualty status, training readiness, personnel, inspections, combat systems, ship readiness and an executive summary. It provides an excellent basis for developing training readiness indicators at the Joint level.

We propose that the training readiness data and data closely related to them, such as personnel and equipment status, be analyzed on the basis of a production model, expressed earlier as input-process-output or a resources-to-readiness paradigm:

1. Analyze data to identify short term and long term trends, including noise, i.e., short term, nonsignificant variations.
2. Where trends are observed, identify the time delays between inputs, i.e., resources, process, and outputs—the related consequences in OPTEMPO and demonstrated combat capability. An important by-product of this examination would be to improve our ability to identify indicators of current and future readiness.

3. Examine indicators for redundancy, i.e., identify those that tend to vary consistently with each other and, thereby, add little additional information about status and trends and that are candidates for elimination.

4. Examine indicators that could be combined by appropriate statistical procedures, perhaps increasing the reliability of the information and reducing the number of indicators to which senior decision makers must attend.

5. Examine the relation between subjective and objective indicators of readiness in an effort to identify the extent to which both are needed and whether the subjective assessments provide information not otherwise available. Extend current overall assessments to include structured evaluations on specified topics.

6. Start the collection and analysis of new demonstrated performance measures such as percent of crews qualified, percent of Training Readiness Examinations above average, percent of Operational Readiness Implications rated excellent or outstanding, and percent of mission essential tasks trained to standard.

REFERENCES

Beland, Russell W., and Aline O. Quester
Interfaces 21, pp. 111-120, July 1991

Burba, Edwin H., Andrew Chambers, Lawrence C. Chambers, Alfred M. Gray, Joseph Metcalf III, Robert D. Russ, and Clyde L. Vermilyea
Training Readiness in the Department of Defense
IDA Document D-1517
Institute for Defense Analyses, Alexandria, VA
May 1994

Cavalluzzo, Linda C.
OPTEMPO and Training Effectiveness
Professional Paper 427
Center for Naval Analyses, Alexandria, VA 22302
December 1984

Cavalluzzo, Linda C.
Unit Training Effectiveness: PACFLT Trip Report
Memorandum for the Record 85-1347
Center for Naval Analyses, Alexandria, VA 22302
17 July 1985

Cavalluzzo, Linda C., and C. Goodwyn
User's Guide to Ship Employment Histories
CRM 85-75
Center for Naval Analyses, Alexandria, VA 22302
September 1985

Cavalluzzo, Linda C., and Stanley A. Horowitz
Resources-To-Ship Training Readiness Study
Final Report, Main Text
Resource Consultants, Inc.
Vienna, VA
July 1987

Cedel, Lt Col, USAF, Thomas E., and Lt Col Ronald P. Fuchs, USAF
An Analysis of Factors Affecting Pilot Performance
Air Force Center for Studies and Analysis, The Pentagon, Washington, DC
December 1986 (AD B109 757L)

Chairman of the Joint Chiefs of Staff
Joint Reporting Structure
Status of Resources and Training System (SORTS)
Joint Pub 1-03.3
Washington, DC
10 August 1993

Chairman of the Joint Chiefs of Staff
Status of Readiness and Training System (SORTS)
Memorandum of Policy No. 11, 1st revision
December 1992

Congressional Budget Office
Trends in Selected Indicators of Military Readiness, 1980 through 1993
Washington, DC
March 1994

Defense Science Board
Report of the Defense Science Board Task Force on Readiness
Office of the Under Secretary of Defense for Acquisition and Technology
Washington, DC 20301-3140
June 1994 (AD 286 412)

Department of Defense
Military Manpower Training Report, FY 1997
Washington, DC
June 1996

Eisenhower, Dwight D.
Memorandum to Admiral Chester W. Nimitz, 12 April 1946
Quoted in "Report of Self-Study: Intermediate-Level Phase II"
Armed Forces College, Norfolk, VA
June 1994

Finch, Louis C.
Analysis for Support of Readiness and Training
Presentation by the Deputy Under Secretary of Defense (Readiness)
to the Military Operations Research Society
Fort Leavenworth, KS
20 June 1996

Follman, D., A. Marcus, and L. Cavalluzzo
OPTEMPO and Ship Readiness
CRM86-123
Center for Naval Analyses, Alexandria, VA 22302
June 1986

Gilman, Harry J., Colin P. Hammon, Stanley A. Horowitz, Jesse Orlansky, and Alec Salerno
Flying Hours and Bombing Accuracy of Aircraft During the Gulf War (U)
IDA Paper P-3299
Institute for Defense Analyses, Alexandria, VA 22311
January 1997 (to be published), SECRET*

Gorman, Paul F.
The Military Value of Training
IDA Paper P-2515
Institute for Defense Analyses, Alexandria, VA 22311
December 1990

Hammon, Colin P., and Stanley A. Horowitz
Flying Hours and Aircrew Performance
IDA Paper P-2347
Institute for Defense Analyses, Alexandria, VA 22311
March 1990

Hammon, Colin P., and Stanley A. Horowitz
Relating Flying Hours to Aircrew Performance: Evidence for Attack and Transport Mission
IDA Paper P-2609
Institute for Defense Analyses, Alexandria, VA 22311
June 1992

Hammon, Colin P., and Stanley A. Horowitz
The Relationship Between Training and Unit Performance for Naval Patrol Aircraft—Revised
IDA Paper P-3139 (Revised)
Institute for Defense Analyses, Alexandria, VA 22311
December 1996

Hiller, Jack H., Howard McFann, and Major General Lawrence Lehowicz, USA
Does OPTEMPO Increase Unit Readiness, p. 71–79
in Holz, Hiller and McFann, editors
Determinants of Effective Unit Performance: Research on Measuring and Managing Unit Training Readiness
U.S. Army Research Institute for Behavioral and Social Sciences, Alexandria, VA
July 1994

---

* No classified information from this source cited in this paper.

Holz, Robert F., Francis O'Hara, and Ward Keesling
Determinants of Effective Unit Performance at the National Training Center: Project Overview
in Holz, Hiller, and McFann, editors
Determinants of Effective Unit Performance: Research on Measuring and Managing Unit Training Readiness, p. 81-96
U.S. Army Research Institute for Behavioral and Social Sciences, Alexandria, VA
July 1994

Horowitz, Stanley A,. and Norma J. Hibbs
Relating Resources to Readiness
CNRI
Center for Naval Analyses
Alexandria, Virginia
September 1979

Horowitz, Stanley, Jesse Orlansky, John C. F. Tillson, Theophilos C. Gemelas, Harry J. Gilman, Colin Hammon, and H. Marshall Hoyler
Unit Training in the Gulf War
IDA Paper P-3087
Institute for Defense Analyses, Alexandria, VA
December 1995

Jareb, Anthony M., Richard A. Heaton, Major James M. Doll, USMC, and Captain J.G. Smith, USMC
Ground Training Readiness Study
CRM 94-131
Center for Naval Analyses, Alexandria, VA 22302
October 1994

Junor, Laura J., and Jessica S. Oi
A New Approach to Modeling Ship Readiness
CRM 95-239
Center for Naval Analyses, Alexandria, VA 22302
April 1996 FOR OFFICIAL USE ONLY

Marcus, Alan J.
Enlisted Crew Turnover and Ship Readiness: Review, Refinements, and Recommendations
CRM 89-169
Center for Naval Analyses, Alexandria, VA 22302
July 1989

Newett, Sandra L., William W. Davis, Douglas S. Esser, Ryan M. Schneider, and Eric S. Miller
F/A-18 Pilot Proficiency Versus Training Resources: Final report (U)
CRM 91-65
Center for Naval Analyses, Alexandria, VA 22302
September 1991, CONFIDENTIAL*

Orlansky, Jesse, and Jack Thorpe, editors
Proceedings of Conference on 73 Easting: Lessons From Desert Storm via Advanced Simulation Technology
IDA Document D-1110
Institute for Defense Analyses, Alexandria, VA
April 1992 (AD A253 991)

Orlansky, Jesse, Carl J. Dahlman, Colin P. Hammon, John Metzko, Henry L. Taylor, and Christine Youngblut
The Value of Simulation for Training
IDA Paper P-2982
Institute for Defense Analyses, Alexandria, VA
September 1994

Prettol, Donald C., Stephen L. Lieberman, Christopher D. Mays, and Thomas G. Godfrey
Collective Unit Training: An Examination of the Training Resource Allocation Process
FP 206 MR1
Logistics Management Institute
McLean, VA 22102
November 1995

Robinson, Matthew T., James M. Jondrow, and Christopher E. Wheeler
Hollowness and the Navy: Interviews with Senior Naval Leaders
CRM 94-167
Center for Naval Analyses, Alexandria, VA 22302
November 1994 (AD B197 965)

Robinson, Matthew T., James M. Jondrow, Laura J. Junor, and Jessica S. Oi
Avoiding a hollow force: An Examination of Navy Readiness
CRM 95-238
Center for Naval Analyses, Alexandria, VA 22302
April 1996 FOR OFFICIAL USE ONLY

---

* No classified information from this source cited in this paper.

Tillson, John C.F.
Building a Joint Training Readiness Reporting System
IDA Paper P-3140
Institute for Defense Analyses, Alexandria, VA 22311
May 1996

Tillson, John C.F., and Steven L. Canby
Alternative Approaches to Organizing, Training and Assessing Army and Marine Corps Units, Part I. The Active Component
IDA Paper P-2791
Institute for Defense Analyses, Alexandria, VA
November 1992

U.S. General Accounting Office
Army Training: Evaluations of Units Proficiency Are Not Always Reliable
GAO/NSIAD-91-72
Washington, DC
February 1991

U.S. General Accounting Office
Army Training: One-Third of 1993 and 1994 Budgeted Funds Were Used for Other Purposes
GAO/NSIAD-95-71
Washington, DC
April 1995

U.S. General Accounting Office
Military Readiness: Current Indicators Need to be Expanded for a More Comprehensive Assessment
GAO-NSIAD-94-160
Washington, DC
April 1994

U.S. General Accounting Office
Military Readiness: Data and Trends for January 1990 to March 1995
GAO/NSIAD-96-111BR
Washington, DC
March 1996

U.S. General Accounting Office
Military Readiness: DoD Needs to Develop a More Comprehensive Measurement System
GAO-NSIAD-95-29
Washington, DC
October 27, 1994

U.S. General Accounting Office
Military Readiness: Improved Assessment Measures are Evolving
GAO-NSIAD-95-117
Washington, DC
March 1995

U.S. General Accounting Office
National Guard: Peacetime Training Did Not Adequately Prepare Combat Brigades for Gulf War
GAO-NSIAD-91-263
Washington, DC
September 24, 1991

U.S. General Accounting Office
Operation Desert Storm: Army had Difficulty Providing Adequate Active and Reserve Support Forces
GAO-NSIAD-92-67
Washington, DC
March 10, 1992

U.S. Navy
Type Commander Readiness Management System
Group/Squadron System Release 5.0: SORTS Module User's Manual
(no date)

Wainstein, Leonard
The Relationship of Battle Damage to Unit Combat Performance
IDA Paper P-1903
Institute for Defense Analyses
Alexandria, VA 22311
April 1986

Weiss, Herbert K.
System Analysis Problems of Limited War, in *Annals of Reliability and Maintainability, Volume 5—Achieving System Effectiveness*
5th Annual Reliability and Maintainability Conference
New York, NY
July 18–20, 1966
Conference sponsored by the American Institute of Aeronautics and Astronautics, The Society of Automotive Engineers and The American Society of Mechanical Engineers, New York, American Institute of Aeronautics and Astronautics
1966

GLOSSARY

ACC	Air Combat Command
ACM	Air Combat Maneuvering
ARG	Amphibious Ready Group
ARI	Army Research Institute
ARTEP	Army Readiness Training Evaluation Program
ATG	Afloat Training Groups
ATRIMS	Aviation Training and Readiness Information Management System
BLTM	Battalion Level Training Model
BLUFOR	Friendly Force
BTM	Battalion Training Model
CALL	Center for Army Lessons Learned
CASREPT	Casualty Report
CAX	Combined Arms Exercise
CBO	Congressional Budget Office
CEP	Circular Error Probable
CINC	Commander-in-Chief
CINCLANTFLT	Commander-in-Chief Atlantic Fleet
CINCPACFLT	Commander-in-Chief Pacific Fleet
CJCS	Chairman of the Joint Chiefs of Staff
COHORT	Cohesion, Operational Readiness Training
COMNAVSURFLANT	Commander Navy Surface Atlantic Fleet
COMPEX	Competitive Exercise
COMPTUEX	Composite Training Unit Exercises
CRP	Combat Readiness Percentage
CTC	Combat Training Center
CVBG	Carrier Battle Group
DO	Deputy for Operations
DOC	Designed Operational Capabilities
DoD	Department of Defense
EIS	Executive Information System
FYDP	Future Years Defense Program
GAO	General Accounting Office
GCC	Graduated Combat Capability
IG	Inspector General
ISIC	Immediate Superior in Command

JMETL	Joint Military Essential Task List
JSCP	Joint Strategic Capabilities Plan
JTFEX	Joint Task Force Exercise
JTS	Joint Training System
JULLS	Joint Universal Lessons Learned System
MAJCOM	Major Command
MCCDC	Marine Corps Combat Development Center
MCCRES	Marine Corps Combat Readiness Evaluation System
MCTRSS	Marine Corps Training Readiness Support System
MILES	Multiple Integrated Laser Engagement System
MPS	Mission Performance Standards
NATO	North Atlantic Treaty Organization
NAVFLIRS	Navy Flight Information Reporting System
NTC	National Training Center
O&M	Operations and Maintenance
OPCON	Operational Control
OPFOR	Opposing Force
OPTASKS	Operational Tasks
OPTEMPO	Operating Tempo
ORI	Operational Readiness Inspection
OSD	Office of the Secretary of Defense
PDC	Predeployment Certification
PDOE	Projected Operational Environment
PMA	Primary Mission Area
POE	Projected Operational Environment
POM	Program Objective Memorandum
QAFA	Quality Air Force Assessment
ROC	Required Operational Capability
SAG	Surface Action Group
SATS	Standard Army Training System
SORTS	Status of Resources and Training System
STRAC	Standards in Training Commission
TRE	Training Readiness Examination
TRM	Training Resource Model
TRMS	Type Commander Readiness Management System
TYCOM	Type Commander
UCOFT	Unit Conduct of Fire Trainer
USACOM	United States Atlantic Command
USAFE	United States Air Force Europe

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	March 1997	Final—January 1996–March 1997

4. TITLE AND SUBTITLE
Indicators of Training Readiness

5. FUNDING NUMBERS
DASW01 94 C 0054
T-L2-1371

6. AUTHOR(S)
Jesse Orlansky, Colin P. Hammon, Stanley A. Horowitz

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Institute for Defense Analyses
1801 N. Beauregard St.
Alexandria, VA 22311-1772

8. PERFORMING ORGANIZATION REPORT NUMBER
IDA Paper P-3283

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
OUSD(P&R)
The Pentagon, Room 1C757
Washington, DC 20301

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT ***(Maximum 180 words)***

This paper is concerned with ways of improving the reliability and accuracy of SORTS, the Status of Readiness and Training System used to report the readiness of the Services for combat to senior officials in the Department of Defense. Although SORTS includes some subjective and potentially unreliable information, indicators of the amounts of training conducted, such as number of flying hours and steaming days, are robust and statistically valid predictors of such combat-related capabilities as bombing accuracy, battle deaths in war, and success in air-to-air combat in exercises and in war. The utility of SORTS for reporting Joint and Service readiness can be enhanced by including certain measures already being used by some of the Services, such as percent of crew qualified and percent of operational readiness inspections rated excellent or outstanding.

14. SUBJECT TERMS	15. NUMBER OF PAGES
readiness indicators, SORTS, training readiness, training accomplishment, training achievement, readiness reports	86
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	SAR